

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:25:46 ; Search time 4250.63 Seconds
(without alignments)
11416.811 Million cell updates/sec

Title: US-09-874-390-1
Perfect score: 2319
Sequence: 1 attaactggttctctatt.....ataataataacttaacaga 2319

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank1:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	ID	Description
1	AX045249	94.6	2216	6	AX045249 Sequence
2	AX045251	94.1	2217	6	AX045251 Sequence
3	AX045253	94.1	2217	6	AX045253 Sequence
4	BC017032	89.7	2106	9	BC017032 Homo sapi
5	AR136078	85.8	2128	6	AR136078 Sequence
6	AF102542	85.8	2128	9	AF102542 Homo sapi
7	AC092755	81.2	187275	2	AC092755 Homo sapi
8	AC092755	81.2	187275	6	AC092755 Homo sapi
9	AF038650	56.8	1317	6	AF038650 Homo sapi
10	AF318573	56.8	1317	14	AF318573 Homo sapi
11	AF231105	43.6	108873	14	AF231105 Homo sapi
12	BC018297	41.4	4179	14	BC018297 Mus muscu
13	BC018297	39.8	1841	10	BC018297 Mus muscu
14	AC096432	31.3	176483	2	AC096432 Rattus no
15	MMU19265	16.2	2033	10	MMU19265 Mus muscu
16	D87333	16.2	2023	10	D87333 Mus muscu
17	D87332	16.1	2047	10	D87332 Mus muscu
18	AX341786	16.1	377	6	AX341786 Sequence
19	AX087946	15.9	1287	6	AX087946 Sequence
20	AX087946	15.9	2105	6	AX087946 Sequence
21	T61440	15.9	2105	6	T61440 Sequence
22	T61440	15.9	2105	6	T61440 Sequence
23	AX333255	15.9	2110	6	AX333255 Sequence
24	AX335500	15.9	2110	6	AX335500 Sequence
25	HUMGOGG	15.9	2110	6	HUMGOGG Sequence
26	HUMGOGG	15.9	2204	9	HUMGOGG Sequence
27	AL138757	15.9	159930	2	AL138757 Homo sapi
28	AL161626	15.9	183858	9	AL161626 Homo sapi
29	S79797	15.7	5010	10	S79797 enzymatic g
30	BT041320	15.7	2435	4	BT041320 Bos tauru
31	AX341869	13.2	309	6	AX341869 Sequence
32	AX340693	13.2	309	6	AX340693 Sequence
33	HSBROVHEZF	11.1	380	14	HSBROVHEZF Sequence
34	AC092754	10.6	174323	2	AC092754 Homo sapi
35	AC092754	10.5	3083	2	AC092754 Homo sapi
36	HSB16NNGT	10.5	151136	2	HSB16NNGT Homo sapi
37	HSB16NNGT	9.0	189534	2	HSB16NNGT Homo sapi
38	AC095187	8.8	252504	2	AC095187 Rattus no
39	AL358777	8.6	153359	9	AL358777 Homo sapi
40	AL358777	8.6	153359	9	AL358777 Homo sapi
41	AC021626	8.3	1362	6	AC021626 Homo sapi
42	AF132035	8.3	3435	9	AF132035 Sequence
43	AB037596	8.2	3508	10	AB037596 Homo sapi
44	AC093259	8.1	184590	2	AC093259 Homo sapi
45	AC087663	8.0	63853	2	AC087663 Homo sapi

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	FEATURES
1	AX045249	Sequence 1 from Patent WO0066727.	AX045249	1	GI:11343799	human.	Homo sapiens	Adolf, G., Helder, K. H. and Sommergruber, W.	Patent: WO 0066727-A 1 09-NOV-2000;	Boehringer Ingelheim International GmbH (DE)	Location/Qualifiers
											1..2216
											/organism="Homo sapiens"
											/db_xref="taxon:9606"
											1..426
											427..1743
											/note="unnamed protein product"

AX045249 Sequence
AX045251 Sequence
AX045253 Sequence
BC017032 Homo sapi
AR136078 Sequence
AF102542 Homo sapi
AC092755 Homo sapi
AC092755 Homo sapi
AF038650 Homo sapi
AF318573 Homo sapi
AF231105 Homo sapi
BC018297 Mus muscu
BC018297 Mus muscu
AC096432 Rattus no
MMU19265 Mus muscu
D87333 Mus muscu
D87332 Mus muscu
AX341786 Sequence
AX087946 Sequence
AX087946 Sequence
T61440 Sequence
T61440 Sequence
AX333255 Sequence
AX335500 Sequence
HUMGOGG Sequence
HUMGOGG Sequence
AL138757 Homo sapi
AL161626 Homo sapi
S79797 enzymatic g
BT041320 Bos tauru
AX341869 Sequence
AX340693 Sequence
HSBROVHEZF Sequence
AC092754 Homo sapi
AC092754 Homo sapi
L42230 Homo sapien
AL121893 Human DNA
AC095187 Rattus no
AC073815 Mus muscu
AL358777 Human DNA
AC021626 Homo sapi
AX087935 Sequence
AF132035 Sequence
AF132035 Homo sapi
AB037596 Mus muscu
AC093259 Homo sapi
AC087663 Homo sapi

	3' UTR BASE COUNT ORIGIN	600 a	499 c	547 g	570 t	
Query Match	Best Local Similarity	94.6%	Score 2194.2	DB 6	Length 2216	
Matches 2196	Conservative	0	Mismatches	3	Indels	0
					Gaps	0
QY	70	gtcacgggaactgacctgctactgttgaaactgacctgaccttaactagcaagttttgttctgg	129			
Db	1	gtcACGGGGAACGCCCTTGCTACTGTGTGACCCGCCCTTACTACACAGTTTGTTCGCG	60			
QY	130	gaagccctgggagttctgctcaactaccatcaactgtagtggctgaaaggaaacgaagaa	189			
Db	61	GAAGCCCTGGGAGTTGTGTAATACCTATACGTAGTGGCTGAAGGAACGATGAAGA	120			
QY	190	acatgacctcaagagagcttccctgctcaatagaagaaccaaagctgagcctgcaagatat	249			
Db	121	ACATGACCTCAAGAGAGCTTCCTGTCATAGAGAGAACCAAGCTGAGCGCTGGCAAGATAT	180			
QY	230	taagaaggaagcctgaaactgttcccttggaacatctagaatgtcagaanaatacctttgg	309			
Db	181	TAAAGAGGAGCGCTGAACGTGTCTCTTGAGACATCTTATGAATGTCAGAAATACCTTTGG	240			
QY	310	agggttagaagaagtcaggggagaaagttgtttcaacattgtctgcacaggaacccggca	369			
Db	241	AGGGTTAGAGATCAGGGGACATGGTGTGTCACATTTGCTGCGACGGAACACCGCAGTC	300			
QY	370	ttcacttggaaacagaatcaacgcctctgtaagaagatacaccctaagcaggaagagcta	429			
Db	301	TTCACTTGGGAACAAATATCACGCCCTTGTGTAAGATCATCCTTAAGCAGAGAGAACTA	360			
QY	430	ctaaagatgtgttccctccacccttccctgtgtctgctggtctcaacgtgttcccatct	489			
Db	361	CTAAAGGATGTGTACTCTCTCACCTTCCTGCTGCGTTCGAGTCCACGTCGCCATTC	420			
QY	450	gtgacgaatggttcaatagtgaagaagactctgcacgctgcacattactgttggctctgg	549			
Db	421	GTGACGATGTGTCATAGGAAGAGACTCTGCCAGCTGATTACTTGTGGGCTCTGGGCTGC	480			
QY	550	tatagtcgtctggcacgtgtgcttgaacctttcttcaagttgaagtgtgaacctgac	609			
Db	481	TATATGCTGTGCGCCACATGTGGCTGTGAACCTTTTTCAGGTTGAAGTGTACTCTGAC	540			
QY	610	caattgggtctgtagtctcaagggatctcaaaagccagtaactgtaagaaatacttgtataat	669			
Db	541	CACTTGGGCTGTGAGTCCAGGGGAATCTCAAAAGCAGTACTGAGGAATATCTTTATAT	600			
QY	670	ttctcgaaacttccagcaagaagatgctatcaactgttcaaggggtcaaccgagggacca	729			
Db	601	TTTCTGAAACTTCCAGCAAAAGAGTCTATCAACTTTTTCAGGGGTCAACCGAGGACAA	660			
QY	730	gaaggaagtctcaagcctatctgaataacctggaggttcaagaagaagcgaaagccttc	789			
Db	661	GAGGAGTGCTTCAAGCTATTCTGATTAACCTGGAGGTCGTAAGAAAGAGAGAGCTTTC	720			
QY	790	acagacaccacatccttccctcaacagagagacttgagcaacttcaagctgtaaaaggag	849			
Db	721	ACAGACACCACTACTCTCCCTTACCAAGAGACTGTGAGCACTTCAAGGCTTAAAGGAAG	780			
QY	850	ttcatacagtttccacgtagcaagaagaaggttggaagttccctattgcatactctatggtg	909			

D	b	781	TTCAATACAGTCCACACTGACCAAGAAGAGTGGATTCCTCATTTGCATACATCTATGGTG	840
Q	y	910	atcaatgagaagalttgaanaactttgaaggtactctgtagctgtgtatgtccctcaagac	969
D	b	841	ATTCAATGAGAAGATTGAAAACCTTTGAAGAGCTACTCCGAGCTGTGTATGGCCCCATGAAAC	900
Q	y	970	atactactgtgtccatgtgtgtgtggaaggtccccaagaactttaaagagcggttaagaac	1029
D	b	901	ATAATACCTGTGCATATGGATGTGAAGATGCCCAAGAACTTTTAAAGAGCGGTCTAAACCA	960
Q	y	1030	attattctctgtctcccaaatgtcttcataatgacgaagaactgtgtctcgagtgttatctc	1089
D	b	961	ATTATTCTTCTCTCCCAATGTCTTCATATGACAGTAACTGGTTCGGTGGTTATATCC	1020
Q	y	1090	tcctgtgtccagggtgtgcaagcttgacactcaatctgcatgaaagaactgtctccaagctcaagt	1149
D	b	1021	TCCTGTGTCAGAGGTGCAAGCTGACATCAATGCATGTGAAGACTGTCTCCACAGCTCAAGTG	1080
Q	y	1150	ccgtgtggaataactcctcgtgaatacatgttggagaagacttctctataagaagcaatgcaag	1209
D	b	1081	CCGTGGAAATATCTTCCATCGAATACATGTGGACGCACTTTCATATAAGCATATCCAG	1140
Q	y	1210	atgtgtccaggtctccagaatgttctgaatggagaagaatgacatgagttcaagaagttaacctct	1269
D	b	1141	ATGTGTCCAGGCTCTCAAGATGTTGATGGAGAAATAGCATGAGTCAAGGTACTCTCT	1200
Q	y	1270	aagcacaagaanaaccgctgtgaaataatcacattgaagtgtgtgtagaagacacatcaccta	1329
D	b	1201	AAGCACAAGAAACCCGCTGGAATATACATTGTAGGTATGTAGAGACACATTAACACTTA	1260
Q	y	1330	accacaagaagaagatctctcccccctataatlttaactaigtlttaacaggaatgtgtac	1389
D	b	1261	ACCAACAAGAAAGAGATCTCTCCCTTAATAATTAATCATGTATTACAGGGATATCCGATC	1320
Q	y	1390	atgtgtgtctccgaagattctgtccacaatgttttgaagaacccttaactcccaacaagt	1449
D	b	1321	ATTGTGGCTTCCGAGATTTTCGTCCACATGTTTATMAAACCCTTAATCCCAACACTG	1380
Q	y	1450	attgaatggtgtgaagaagacacttataagcccaatgaaacacctcttggccaaccttcaagct	1509
D	b	1381	ATTGAATGGGTAATAAAGACACTTTATAGCCCAAGATGAAGACACTCTGGGGCACCCCTTACGGT	1440
Q	y	1510	gcaaggtgatagtcgtgtgtctgttcccaaccaccacaagttacgacacatctcaagaatgact	1569
D	b	1441	GCACGCTGGATGCTCTGGCTCTGTTCCTCAACACCCCAATAGACATCTCAACATATGACT	1500
Q	y	1570	tctattgtccaggtcgtgtcgaatgttgacaggtgtcatagaaaggaatcgaatlaaaggtgtctct	1629
D	b	1501	TCTATTGTCAGAGCTGTGTCAATGTGGCAGGGTCATGAGGGAACATGATGAAGGTCTCTCT	1560
Q	y	1630	tatgtctccctgtctcttgaatcccaacagcgaggtcatctgttlatctgtgtgtgtgtgtgtgt	1689
D	b	1561	TATGTCTCCCTGTGTGAATCCACACAGCGGGCATATGTGGTTATGTGGGCTGGGGCACTTG	1620
Q	y	1690	aattgtgatbctcaaaaacacatcacctgttggccaacaagtttgaccaaaaggtatagat	1749
D	b	1621	AATTGGATGCTTCAAAACCATACCTGTTGGCCAAACAAATTGACCCCAAAAGGTATGATAT	1680
Q	y	1750	aatgtctctcaagtgtcttagaagaataactacgttataaagccaattataggaaactgaact	1809
D	b	1681	AATGTCTTCACTGCTTAAAGAAATATCATGTTAATAAGCCATCTATGTGGACTGTAACTT	1740
Q	y	1810	tgagaacaacatataagagcgttgcctaccgttgggtgcaaaagcatgtacaacaatgtctag	1869
D	b	1741	TGAGACACACATATAGAGCTGTGCTACTGTGGGGCAAGACATGTACAAACATCTCAG	1800
Q	y	1870	aactgtctgtgagacagtgtgtgtgtgtgagaacccaaggtcttjcaatctgtgtgacatccttag	1929
D	b	1801	AACTTGCTGGGACAGTGTGTGGTGGAGACACAGGGCTTTCAATTTGTGTGCATCTTTAGG	1860
Q	y	1930	ataagagggtctgctataagattgtgtgtgtatgatatcttctgtccttgcaaatgtgtgct	1989

Db 1861 ATAAAGAGGCTGCTATTAGATTGTGGTAAAGATCTTTTCCCTGCAAAATTCGCGCT 1920
QY 1990 gggtaatgctgctgtctctccaccctaaacctagtagttccctccactaacttctca 2049
Db 1921 GGGTAAATGCTCTGTTCTCCACCCTTAACCCCTAGTAGTTCCCTCACTTAATCTTCTCA 1980
QY 2050 ctaagtgaatgagaaactcgtctatagggagagtgaaaggagatagtgtagagca 2109
Db 1981 CTAATGAGAAATGAGAACTGCTGTATAGGAGAGTGAAGAGGAGATATGTGTAGAGCA 2040
QY 2110 ctgattcagttgtaagctcgtcgtgtagctttccatctctgtgagctgcgtctctaa 2169
Db 2041 CTGTGATTCAGTTGATGAGCTGCTGTAGCTTTTCATTCCTGTGAGCTGCCGTCTCA 2100
QY 2170 taattccaggttggtagcgtgtgagagaaacttgatgaaagaaaccttccctctgt 2229
Db 2101 TAAATCCAGGTTGGTAGGAGGAGAACTTGAAGAAAGAAACCTTCCCTTCTGT 2160
QY 2230 actgttaacttaaaaaataatagctccctgattcaagta 2268
Db 2161 ACTGTTAACTTAAATAATAGCTCCGTGATCAAGTA 2199

RESULT 2
AX045251 2217 bp DNA linear PAT 24-NOV-2000
LOCUS AX045251
DEFINITION Sequence 3 from Patent WO0066727.
ACCESSION AX045251
VERSION AX045251.1 GI:11343801
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2217)
AUTHORS Adolif, G., Heider, K. H. and Sommergruber, W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 3 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
source location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
5'UTR 1..844
5'UTR 1..426
CDS 427..999
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17335.1"
/db_xref="GI:11343802"
/translation="MVOMKRLCOLHYLMALGCVMLIATVALKLSFRLKCDSDHGLIES
RESQSYCRNLTYFPLKIPAKMSINCSVTRGDDEAYLQALINLKYKKKEPTDTH
YLSLIRDCENHFKARKFTQPLSKEEVEFPLAYSMVIEHKEINERILLRAYADONIT
CVHVDVEPRNFQSGSYNFFLPRKLSQ"
845..1744
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17336.1"
/db_xref="GI:11343803"
/translation="MRRILKGLKCELMPLRTYVSMKMKSPETKEAVKATISCF
PNVEIASLIVVYVYASRVQADLNCMEDLLQSSVPMKYFLNTGTFEPISNEMO
ALKMLNGRNSMESEVPPKHKETRKHYHEVVRDILHLJNKKKDDPPYNIJMTGNAYI
VASRDPVOHVLKMKNSQQLILEMVDYTSPEDEHMLATIQARAKMFGSPVNHKXIDSM
TSTARLVKMGHESDIDKGIAPLPCSGIHOAICVYAGDLMMLONHHLIANKFDPK
VDNLAQCLELEYKAIYGTETL"
1000..2217
3'UTR 1745..2217
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match 94.1%; Score 2183.2; DB 6; Length 2217;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 2196; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 70 gtcaagggaagcgcctctgtctacttctgtgaccccttaccagagtttctgtctg 129
Db 1 gtcaagggaagcgcctctgtctacttctgtgaccccttaccagagtttctgtctg 60
QY 130 gaagccctggattctgtcaataacatcactgtatgtgtcgaaggaacagatgaaga 189
Db 61 GAAGCCCTGGATTCTTGCTAATACCTATCTAGTACTCTAAGAGAAACAGATGAGA 120
QY 190 acatgaacctcaaggagcttcctgtcaatgaagaagcaagctgaagctgtgcaaatat 249
Db 121 ACATGACCTCAAGAGAGCTTCTGTCAATGAGAAGCAAGCTGAGCGCTGCCAAGATAT 180
QY 250 taaagaaggccttgaaactgttctcttgagacatctatgaatgtcagaataactcttgg 309
Db 181 TAAAGAGGAGCTGAAACGTTCTTGACACTTATGATGATGAGAAATACCTTTTGG 240
QY 310 aggttaagaagatcaaggagacatgtgtctacattgtgtcagaggaacacgcgaatc 369
Db 241 AGGTTAAGAAATCAAGGGGACATGGTTGTTCACATTTGCTGCCACGGAACACGCCAGTC 300
QY 370 ttcacttggaaacagaatcaagccttgtgaagagatcaatccctaaagcaggagaagata 429
Db 301 TTCACCTGGGAACAGAAATCACGCCCTTGTGAAGAGATCACTTAAGCAGAGAGAAAGTA 360
QY 430 ctaaggatgtgtctctccctccacttccctgtgtctgtccacactgtctccacttct 489
Db 361 CTAAAGGATGTGTACTCTCTCCACTCTCCCTGTGCTGGTGTCCACCTGTCTCCATTTT 420
QY 490 gtgaagatgttcaacttgaagaagaccttcagcagctgcatattctgtggtctgtggtc 549
Db 421 GAGAGATGCTTCATGAGAGAGACTGTCCACGCTGCATATTGTGGGCTGTGGGCTGC 480
QY 550 tatatgtctgtgccaactgtgtctgtgaaacttcttcaagtlgaagtgtgactgtgac 609
Db 481 TATATGCTGCTGGCCACTGTGCTGTGAATCTTCTTACGTTGAAGTGAAGTCACTTGAC 540
QY 610 cacttgggtctggaatccaggaatctaaagcagctctgtgaagaatatctgtatata 669
Db 541 CACTTGGGTCTGAGTCCAGGGAATCTCAAGCAAGTCAAGTCAAGTAATCTTGTATATAT 600
QY 670 ttcttgaacacttcagcaagaaggtctatcaactgttccaggggtcaaccgaggagca 729
Db 601 TTCTGGAACCTTCAGCAAGGTGCTATCACTGTTCAGGGGTCAACCGAGGAGCA 660
QY 730 gaggcagtgctcaagctattctgaataacctggaagtgtaagaagaagcagaccttct 789
Db 661 GAGCAGTGTCTCAGGCTATTCTGAATTAACCTGAGAGTCAAGAAAGCAGAGCCTTTC 720
QY 790 acaagaccccaactaccttctccctaccagagactgtgagcaatccaaggtctgaagaag 849
Db 721 ACAGACACCACTACTCTCTCCCTCACAGAGACTGTGACACTTCAAGGCTGAAGGAAG 780
QY 850 ttcatcaagttcccaactgagcaagaagaggtgtgaggttccctattgcaactctatgtg 909
Db 781 TTCTATACAGTTTCCCACTGAGCAAAAGAGGTGGAGTTCCCTATTTGCTATGTGGTG 840
QY 910 altcatgaagaatltgaaaactltgaaggctactgtgagctgtgtatgtgcccccaagac 969
Db 841 ATTATGAGGAAGATTGAAGAACTTTGAAGAGGTACTGCGAGCTGTATGCCCTCAGAAC 900
QY 970 atatactgtgtccatgtgtgagtg-agaagtcctccagaagaacttccaagggtcggtcaagc 1028
Db 901 ATATCTGTGCTCAATGTGATGAAAGTCCCAAGAACTTTCAAGAGGCGGTCAAGC 960
QY 1029 aattattcttgcctcccaaatgtctcatagccagtaagctgtgtgtgtgttctatgc 1088
Db 961 AATTATTTCTTGTCTTCCCAATGTCTTATAGCCAGTAAGCTGTGTGGGTGTTATATGC 1020
QY 1089 ctctgtgtccaggtgtgcaagctgtaacctcaactgtcaatgtgaagactgtctcagaagctcagt 1148
Db 1021 CTCTGTGTCCAGGGGTGCAAGCTGACCTCAACTGATGGAAGACTTGTCTCAGAGCTCAGT 1080

QY 1149 gccctggaatcactccctgaatcatgctggagcgaacttcctataaagacaatgcaga 1208
 |||||
 Db 1081 gccgtggaatcactccctgaatcatgctggagcgaacttcctataaagacaatgcaga 1140
 QY 1209 gatgctccggcctcctaagaatgttgaaatggagaataagatgagctcagaagtactcc 1268
 |||||
 Db 1141 gatgctccggcctcctaagaatgttgaaatggagaataagatgagctcagaagtactcc 1200
 QY 1269 taagacaagaagaccgcctggaataatacacttgagtgatgagacacatcacct 1328
 |||||
 Db 1201 taagacaagaagaccgcctggaataatacacttgagtgatgagacacatcacct 1260
 QY 1329 aaccaacaagaagaagatccctccctataatctaatactatgttcaaggaaatgccta 1388
 |||||
 Db 1261 aaccaacaagaagaagatccctccctataatctaatactatgttcaaggaaatgccta 1320
 QY 1389 catgtggtctcccgagatctgtccaacatgtttgaaagaaccctaataccacaact 1448
 |||||
 Db 1321 catgtggtctcccgagatctgtccaacatgtttgaaagaaccctaataccacaact 1380
 QY 1449 gatgaaatgggtaaagaacactatagaccagaatgaacactctgggccaaccttgaag 1508
 |||||
 Db 1381 gatgaaatgggtaaagaacactatagaccagaatgaacactctgggccaaccttgaag 1440
 QY 1509 tgcacggtggaatgctgctgtcttcccaacacaccccaagtaacacatcacaacatgac 1568
 |||||
 Db 1441 tgcacggtggaatgctgctgtcttcccaacacaccccaagtaacacatcacaacatgac 1500
 QY 1569 tctcatatggcagctgctgaagtgagcagagctcatgagggagacatcgataagtgctcc 1628
 |||||
 Db 1501 tctcatatggcagctgctgaagtgagcagagctcatgagggagacatcgataagtgctcc 1560
 QY 1629 ttaagtcctcctgtctggaatcccaacagcggtcatctgcttataatggctggagact 1688
 |||||
 Db 1561 ttaagtcctcctgtctggaatcccaacagcggtcatctgcttataatggctggagact 1620
 QY 1689 gaattggaatgctcaaaacacatcactgttggccaacaagtltgaaccacaagtgatga 1748
 |||||
 Db 1621 gaattggaatgctcaaaacacatcactgttggccaacaagtltgaaccacaagtgatga 1680
 QY 1749 taatgctctcactgcttgaagaataatcactaagttataagcaccatctatgagctgaact 1808
 |||||
 Db 1681 taatgctctcactgcttgaagaataatcactaagttataagcaccatctatgagctgaact 1740
 QY 1809 ttgagacacataagaaagcttgcctacacgttgaggcaagacatgatacaacaatgctca 1868
 |||||
 Db 1741 ttgagacacataagaaagcttgcctacacgttgaggcaagacatgatacaacaatgctca 1800
 QY 1869 gaactgctggaacatgtggtgaggagacaaaggtcttgaatctcgatggcatcctttag 1928
 |||||
 Db 1801 gaactgctggaacatgtggtgaggagacaaaggtcttgaatctcgatggcatcctttag 1860
 QY 1929 gataaagagagctcctatgaatgttgaggtaagatccttttccttggaatgtgcgc 1988
 |||||
 Db 1861 gataaagagagctcctatgaatgttgaggtaagatccttttccttggaatgtgcgc 1920
 QY 1989 ttgggtgaatgctgctgtctcaccaccccaacactagtagtccctccacaatactctc 2048
 |||||
 Db 1921 ttgggtgaatgctgctgtctcaccaccccaacactagtagtccctccacaatactctc 1980
 QY 2049 actaaatgagaatgagaactgctgtgataagagagagtgaagagaggaatgctgtagagc 2108
 |||||
 Db 1981 actaaatgagaatgagaactgctgtgataagagagagtgaagagaggaatgctgtagagc 2040
 QY 2109 acttgatttgaatgaaatgctgtgtgataagcttttttccactctgtagagcgccttcccta 2168
 |||||
 Db 2041 acttgatttgaatgaaatgctgtgtgataagcttttttccactctgtagagcgccttcccta 2100
 QY 2169 ataattccaggttctgtagcgttgagaagaacttgaatgagaagaagaactccctctctg 2228
 |||||
 Db 2101 ataattccaggttctgtagcgttgagaagaacttgaatgagaagaagaactccctctctg 2160

QY 2229 tactgttaactaaataaataagctccctgatacaagta 2268
 |||||
 Db 2161 tactgttaactaaataaataagctccctgatacaagta 2200
 RESULT 3
 AX045253
 LOCUS
 DEFINITION Sequence 5 from Patent WO006727.
 ACCESSION AX045253
 VERSION
 KEYWORDS AX045253.1 GI:11343804
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 (bases 1 to 2217)
 AUTHORS
 Adolf, G., Heider, K.H. and Sommergruber, W.
 TITLE
 Tumour-associated antigen
 JOURNAL
 Patent: WO 006727-A 5 09-NOV-2000;
 Boehringer Ingelheim International GmbH (DE)
 FEATURES
 Location/Qualifiers
 source
 1..2217
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 5'UTR
 1..844
 845..1744
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17337.1"
 /db_xref="GI:11343805"
 /translation="MRRLTKLNGYCELCMPLEFVTVSMNMKSPTEKAVKAIISCF
 PNVFLASKLRYVVAWSRVQADLMCEMDLQSSVPMKFLVTCGTDPPISNABWQ
 ALKMLNGRSMSEVPEPKRETRMKYHEFVVDLTLNKKKDDPPYNIPTGNAIY
 VASRDFVQVILNPKRSOOLIEWYKDTYPSDEHLMFLORARMPQSVNPKYDLSDM
 TSIALVLRKQGHGEBIDKGAIRPQSGIHQRIACVYGACDLNMLQNHILANKRDPK
 VDNALQCLEELRYKAIYTEL"
 1745..2217
 3'UTR
 BASE COUNT 600 a 498 c 547 g 572 t
 ORIGIN
 Query Match 94.1%; Score 2183.2; DB 6; Length 2217;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2196; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 70 gtcaaggaaactccttgacttgtaactgccccttaactcagcaatttctctg 129
 |||||
 Db 1 gtcaaggaaactccttgacttgtaactgccccttaactcagcaatttctctg 60
 QY 130 gaagccctgggaatctgctaatacctacatcagttagtgctgaaggaaacagatgaaga 189
 |||||
 Db 61 gaagccctgggaatctgctaatacctacatcagttagtgctgaaggaaacagatgaaga 120
 QY 190 acatgacctaaagaagacttctgtcaatgagaagacaagctgaagcctggacaagatat 249
 |||||
 Db 121 acatgacctaaagaagacttctgtcaatgagaagacaagctgaagcctggacaagatat 180
 QY 250 taaagaggaagcctgaagactgttccttggaacatcttgaatgcagaataatcacttgg 309
 |||||
 Db 181 taaagaggaagcctgaagactgttccttggaacatcttgaatgcagaataatcacttgg 240
 QY 310 aggttgaagaatcaggaggacatggtgtcacattgtctgcacaggaacacccagcagtc 369
 |||||
 Db 241 aggttgaagaatcaggaggacatggtgtcacattgtctgcacaggaacacccagcagtc 300
 QY 370 ttcaacttgaagaacagatacagccttgtgaagagatcatccctaaagcaggagagaagctca 429
 |||||
 Db 301 ttcaacttgaagaacagatacagccttgtgaagagatcatccctaaagcaggagagaagctca 360
 QY 430 ctaagagatgtgtcctcccaaccttccctgagcaggtgtccacacttcccaatcc 489
 |||||
 Db 361 ctaagagatgtgtcctcccaaccttccctgagcaggtgtccacacttcccaatcc 420

Oy 440 gtgacgatggttcaatggaaagagactctgcccagctgcatctactgttggtgtctctgggtctgc 549
 Db 421 gtgacgatggttcaatggaaagagactctgcccagctgcatctactgttggtgtctctgggtctgc 480
 Oy 550 tataatgctgctggccacgctgtgtcttgaactcttcttcagggtgaagtgtgactctgcac 609
 Db 481 tataatgctgctggccacgctgtgtcttgaactcttcttcagggtgaagtgtgactctgcac 540
 Oy 610 caattgggtctgaggttccagggaactcctcaagccagttactgttgaagaaatctctgtataat 669
 Db 541 caattgggtctgaggttccagggaactcctcaagccagttactgttgaagaaatctctgtataat 600
 Oy 670 ttcttgaaacttccagcaagaagagctcactcaactgttccagggttccagccaggaggagccaa 729
 Db 601 ttcttgaaacttccagcaagaagagctcactcaactgttccagggttccagccaggaggagccaa 660
 Oy 730 gaggcagtgctcaggctacttcttgaaataacctggaaggttcagaagaagcagagaccttc 789
 Db 661 gaggcagtgctcaggctacttcttgaaataacctggaaggttcagaagaagcagagaccttc 720
 Oy 790 aagagacccactactctctccctccagagactgttgagcaactcgaagcttgaagaagag 849
 Db 721 aagagacccactactctctccctccagagactgttgagcaactcgaagcttgaagaagag 780
 Oy 850 ttcaacagcttcccaactgagcaagaagaggtgaggttccctattgcatctactgtgtg 909
 Db 781 ttcaacagcttcccaactgagcaagaagaggtgaggttccctattgcatctactgtgtg 840
 Oy 910 attcaaggaagattgaanaactttgaagagctactgagctgtgtgtatgtgcccccaagaac 969
 Db 841 attcaaggaagattgaanaactttgaagagctactgagctgtgtgtatgtgcccccaagaac 900
 Oy 970 atatactgttccatgtgtgagt-agaagttcccaagaacttccaagaagggcggttcaagc 1028
 Db 901 atatactgttccatgtgtgagt-agaagttcccaagaacttccaagaagggcggttcaagc 960
 Oy 1029 aattaattctgtcttcccaaatgtcttcaatagcagtaagctgtgttgcgtgttctatgc 1088
 Db 961 aattaattctgtcttcccaaatgtcttcaatagcagtaagctgtgttgcgtgttctatgc 1020
 Oy 1089 ctctgtgtccaggtgtgcaagctgaccccaactgcatgtgaagaactgtgtccagagctcagt 1148
 Db 1021 ctctgtgtccaggtgtgcaagctgaccccaactgcatgtgaagaactgtgtccagagctcagt 1080
 Oy 1149 gccgtggaataactctcctgaataactgtgtggaagcagacttctctataagaagagaatgcaga 1208
 Db 1081 gccgtggaataactctcctgaataactgtgtggaagcagacttctctataagaagagaatgcaga 1140
 Oy 1209 gatgttccaggtctcagaagtgtgaatgggaggaatagcatgtgaggtcagaggtaccttc 1268
 Db 1141 gatgttccaggtctcagaagtgtgaatgggaggaatagcatgtgaggtcagaggtaccttc 1200
 Oy 1269 taagcacaagaagaaccgcgtggaatatacactttagaggtagtgtgagagacacatlaacact 1328
 Db 1201 taagcacaagaagaaccgcgtggaatatacactttagaggtagtgtgagagacacatlaacact 1260
 Oy 1329 aaccacaagaagaagaagatccctccctataacttaactaactgtttaaagggaatgcgtta 1388
 Db 1261 aaccacaagaagaagaagatccctccctataacttaactaactgtttaaagggaatgcgtta 1320
 Oy 1389 catgtgtcttcccgagattctgtccaacatgttcttgaagaaccctaaatcccaacaact 1448
 Db 1321 catgtgtcttcccgagattctgtccaacatgttcttgaagaaccctaaatcccaacaact 1380
 Oy 1449 gatgtgaatggttaaaagacacttataagccagatgaacacactctgtggccaaccttcagcg 1508
 Db 1381 gatgtgaatggttaaaagacacttataagccagatgaacacactctgtggccaaccttcagcg 1440
 Oy 1509 tgcagagtgagtgctgtgtctgttcccaacaccccaagtaagcagcatctagaagaatgc 1568
 Db 1441 tgcagagtgagtgctgtgtctgttcccaacaccccaagtaagcagcatctagaagaatgc 1500

Oy 1569 ttctattgccaggtctgtcaaatggtgaggttcatatgagggagacatgcataaagggtgtctc 1628
 Db 1501 ttctattgccaggtctgtcaaatggtgaggttcatatgagggagacatgcataaagggtgtctc 1560
 Oy 1629 ttatgtctccctgtctgtgaatcccaagcgggtctatctgcgtttatgggtctggggact 1688
 Db 1561 ttatgtctccctgtctgtgaatcccaagcgggtctatctgcgtttatgggtctggggact 1620
 Oy 1689 gaattggagttctcaaaaacatcacactgtgttggccaagaagtttgacccaaggtatgata 1748
 Db 1621 gaattggagttctcaaaaacatcacactgtgttggccaagaagtttgacccaaggtatgata 1680
 Oy 1749 taatgtcttctcagttgttgaagaatatacctaagcttataagccatcatatlgaggtgaact 1808
 Db 1681 taatgtcttctcagttgttgaagaatatacctaagcttataagccatcatatlgaggtgaact 1740
 Oy 1809 ttgaagacacactatgagagcgttctactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1868
 Db 1741 ttgaagacacactatgagagcgttctactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1800
 Oy 1869 gaactgtgtgagcaggt 1928
 Db 1801 gaactgtgtgagcaggt 1860
 Oy 1929 gataagaggtctgtcattagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1988
 Db 1861 gataagaggtctgtcattagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1920
 Oy 1989 tgggtgtgaatgtctgt 2048
 Db 1921 tgggtgtgaatgtctgt 1980
 Oy 2049 actaagtgagaatgtgaatgt 2108
 Db 1981 actaagtgagaatgtgaatgt 2040
 Oy 2109 actgattcagttgaatgt 2168
 Db 2041 actgattcagttgaatgt 2100
 Oy 2169 ataattccaggttgt 2228
 Db 2101 ataattccaggttgt 2160
 Oy 2229 taactgttaacttaaaaataaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2288
 Db 2161 taactgttaacttaaaaataaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2200

RESULT 4

LOCUS BC017032 2106 bp mRNA linear PRI 09-NOV-2001
 DEFINITION Homo sapiens, glucosaminyl (N-acetyl) transferase 3, mucin type,
 clone MGC:9086 IMAGE:3851937, mRNA, complete cds.
 ACCESSION BC017032
 VERSION BC017032.1 GI:16877561
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2106)
 AUTHORS Strausberg R.
 TITLE Direct Submission
 JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK
 COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: gcgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdickexil.stanford.edu
Dickson, M., Schmitz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 20 Row: n Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4758421.

FEATURES

source
1. 2106
/organism="Homo sapiens"
/db_xref="locusID:9245"
/db_xref="taxon:9606"
/clone="MGC:9086 IMAGE:3851937"
/tissue="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
309. 1625
/codon_start=1
/product="glucosaminyl (N-acetyl) transferase 3, mucin
type"
/protein_id="AAH17032.1"
/db_xref="GI:16877562"
/translation="MYOMKRLCOLHYALMGCMYLAATVAKLSPFLKDDSHLGLS
RBSQSOYCRHILNPLKPKRSINSQVTRGQEVLAQILNLNLEVKKREFPTTH
YLSLTDCEHFKEKRFIOPPLSKSEVEFPPLASMTHEKTENEFLRLRAVAPONTY
CVAVDEKSPETKEAVKAIISCPNVEIASKLRYVYASRQADLNKMEDLLQSSV
PMKYFNTGCTDPIKSNAMVOALKMLNRNMESEVPRKHETKHYFEVBDTL
HLNKKKDPPEVNLMTGNAYIVASRDFQHYLAKPKSQQLEMYKQDYSPENHMA
TLORARMPGVSYPNHPKXDIDSMKTSIARLYKKMGHGHDIDKGAIPVPCSGIHRALICV
YKAGDLMNMLQNHILANKFDPKVDNALQCELYLRKAIYTEL"

CDS

BASE COUNT 582 a 471 c 516 g 537 t
ORIGIN

Query Match 89.7%; Score 2081; DB 9; Length 2106;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2081; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 gaacatgacctcaagaagctctctgtcaatgagaagcaagaagctgagcctgcaaat 247
DB 1 GAACATGACCTCAAGAGAGCTTCCTGTCATGAGAGAACCAAGCTGACGCCGCAAAAT 60
QY 248 attaaagaagagcctgaaactgtctcttgacatcttaagaatgcagaataactcttt 307
DB 61 ATTAAAGAGAGGCTGAAAGCTTCTTGACATCTTATGATGTCAGAAATACCTTTT 120
QY 308 ggaaggttagaagaatcgaaggacatggtgttcaatctgtctgcacaggaacacccgag 367
DB 121 GGAGGGTTAGAAAGATGAGGACATGCTGTCACATTTGCTGACAGGAAACACCGGACG 180
QY 368 tctccacttgaaacagaatcaacgctctgtaagaagatacctccctaagcagaagaagc 427
DB 181 TCTTCACTTGGAAACAGATACGCCCTTGTGAAGAGATCATCCTTAAGCAGAGAGAAAGC 240
QY 428 tactaaagaattgttctctccctccacttccctgtgtctggtctcgaactgtccact 487
DB 241 TACTAAAGAGATTGTCTCTCTCCCTCCACCTGCTGCTGCGGTCCACACCTGTCCCAT 300
QY 488 ctgtgagaatggttcaatgagaagactctgcagctgcattactgttggctctggct 547
DB 301 CTGTGAGAGATGCTTCATGGAAGAGACTGTCCAGCTGCATTAATTGTGGCTGTGGCT 360
QY 548 gcatatgctgctgcaactgtgcttgaactcttctaagttgaagtgtactctg 607
DB 361 GCTATATGCTGCTGCGCACTGTGCTGTGAACCTTTCTTTCAGGTTGAAGTGTACTCTG 420

QY 608 accacttggctctgagtgccagggaatcacaagccaactctttagaataatctgtata 667
DB 421 ACCACTTGGCTGTGAGATGATCAGAGGAATCTCAAGCAGTACTAGAAATATCTTGATA 480
QY 668 attcctgaacttccagaagaaggtctatcaactggttaagggttccaccggaggacc 727
DB 481 ATTTCTGAAACTTCCAGCAAGAGGCTATCAACTGTTAGGGGTCCCGAGGGAGACC 540
QY 728 aagagcagtgcttcaagctatctgataaacttgagtgagtgcaagaagaagcgaagcct 787
DB 541 AAGAGGAGTGCTTCAGAGCTATCTGAAATAACTGGAGGTCAGAGAAAGACGAGAGCTT 600
QY 788 tcacagaaccacatccttccctcaacgaagactgtgacacttcaagcttgaaga 847
DB 601 TCACAGACACCCACTACCTCTCCCTCCACGAGCTGTGAGCACTTCGAAGCTGAAAGGA 660
QY 848 agttcaacagttccactgagcaagaagaagtgagttccctatgtgactatctatg 907
DB 661 AGTTCATACAGTTCCCACTGAGCAAGAGAGGTGAGTCCCTATTTCATCTATGCG 720
QY 908 tgattcagagaagatgaaacttgaagagctactgcagactgtgtatgtccctcaga 967
DB 721 TGATTCATGAGAAGATTGAAAACTTTGAAGGGCTACTGCGAGCTGTGTATGCCCTCAGA 780
QY 968 acataactgttccatgtgtatgagaagatgcccaagaacttccaagaagcgtgcaag 1027
DB 781 ACATATACCTGTGTCATGATGATGAGAGAGTCCCGAAGAACTTTCAGAGCGGCTCAAG 840
QY 1028 caattattcttcttcccaatgtcttcaatgaacgaatgaagcgtgttcggtgtttatg 1087
DB 841 CAATTATTTTCTGCTTCCCAATGTCTTCATAGCCAGTAAGCTGTGCGGTGTTATG 900
QY 1088 cctctctgctcaggggtgcaagcttgaactcaactgcatgagaagactgtctcagaagctcag 1147
DB 901 CCTCTGCTGACAGGCTCAAGCTGACCTCAATGCAATGCAAGAACTTCTCAGAGCTCAG 960
QY 1148 tgcgctggaataactcttcaatacatgtgtggagcagacttctctataaagaatgag 1207
DB 961 TGCGGTGGAATATCTTCTTAATACATGATGAGAGGAGCTTCTTCAATMAAGCAATGACG 1020
QY 1208 agatgtccagagcctcacaagatgttgaatgagggaatgacatgagtgacac 1267
DB 1021 AGATGTCCAGAGCTCTCAAGATGTTGAATGGAGGAATACATGAGACTCAGAGTACTCTC 1080
QY 1268 ctaagcaacaagaaccgcttggaatatatacttgaagtagtgaagacacattacacc 1327
DB 1081 CTAAGCAACAAGAAACCGCTGGAAATATCATCTTGAGTAGAGACACATTAACAC 1140
QY 1328 taaccacaagaagaagatctccctctataatctaactatggttcaagggaatgct 1387
DB 1141 TAACCAACAAGAAAGAAAGATCTCCCTCTTAATTAATTAATTAATTAATTAATTAATTA 1200
QY 1388 acattgtgcttcccgagatcttgcacaatgttgaagaaccctaactccaacaac 1447
DB 1201 ACATTGTGCTCTCCGAGATTTGCTCCACATGTTTGAAGAACTTAATTCACCAACAC 1260
QY 1448 tgattgaatggttaaaagaactatagccagatgagacaactctgtggccaacctcagc 1507
DB 1261 TGATTGAATGGGTAAGAAAGACATATATAGCCAGATGAACACCTGTGGCCACCTTGACG 1320
QY 1508 gtgcaggtgtagtgcctgtgtctgttcccaacaccccaagtagacatctcaagaatga 1567
DB 1321 GTGCAGGCTGAGGCCCTGCTGTTGTTCCCAACACCCCAAGTATGAGATCTAGACATGA 1380
QY 1568 ctctctatgccaagctgtgtaagtggaggttatalagggaagacatgataaagggtgtc 1627
DB 1381 CTCTATTTGCAAGGCTGTGTAAGTGCAAGGCTCATGAGGAGACATCGATAAGGGTCTC 1440
QY 1628 ctatgtctccctgctctggaatccaccagcggctactctgttlatggtgtggagact 1687
DB 1441 CTATATGCTCCCTGCTGTCGGAATCCACAGCGGGGTATCTCGTTTATGGGGCTGGGACT 1500
QY 1688 tgaattgagatgtctcaaacaccatcaactgttggccaagaagttgaccacaaggtatg 1747

|||||
Db 1501 TGAATGATGCTTCAAAACCATCACCTGTGGCCAAACAGTTTACCCCAAGATGAGTG 1560
Qy 1748 ataatgctcttcagtcttaagaagaatacctaagctataagccaataatgagactaac 1807
Db 1561 ATAATGCTCTTCAGTGGCTTAAGAAATACCTACGTTATAGGCCATCTATGGAGCTGAAC 1620
Qy 1808 tttagaacaacataagaaagctgtctacccctgtggggaagaagacatgacaacatgttc 1867
Db 1621 TTGAGACACACTATGAAAGGCTGCTACCTGTGGGCAAGAGCAATGACAAACATGCTC 1680
Qy 1868 agaactgtcgtggagacgtgtgtgtggagacaagagcttgcacatcgttgagatccctta 1927
Db 1681 AGAATCTCTGGGACAGTGTGGGTGGAGACAGGCGCTTGGAAATGCTGGATCTTTTA 1740
Qy 1928 ggataaagagagctgctataagatgtgtgtgtgaagatcctttgccttgcaaatgtctgc 1987
Db 1741 GGATTAAGAGGCTGCTATTATATGCTGGTAAGTATCTTTGGCTTGCAAAATGCTGC 1800
Qy 1988 ctggagtgatgtctgtctgtctcaccaccctaaacctagtgatctcactaaccttct 2047
Db 1801 CTGGGTGAATGCTGCTGTTTCTCTCTACCCCTAACCCCTAGTATCTCTCACAATCTTCT 1860
Qy 2048 caactaagtgaagaatgaagaactgtgtgtgaaggaagagtgaaagaggaatagtgtgtagag 2107
Db 1861 CACTAAGTGAATGAGAACTGCTGTGATAGGAGAGAGTGAAGAGAGATATGTGTAGAG 1920
Qy 2108 caactgattcagtgtgaatgctcgtcgtgtgaactttccatctgtgtgagctgcgttcc 2167
Db 1921 CACTGATTTTCAATGTAATGCTGCTGCTGATCTTTCATCTGTGAGCTGCCCTTCT 1980
Qy 2168 aataatccaggtttgtgagctgtgaggaagaaacttgaatgaaagaagaaacttccctct 2227
Db 1981 AATAATTCAGAGTTTGGTATGCGGTGAGAGAGAACTTTGATGCAAGAGAACTTCCCTTCT 2040
Qy 2228 gtactgttaacttaaaataaataagctcctgaattcaaaaga 2268
Db 2041 GTACTGTAACTTAATAATAATAGCTCTGATTCAAAGTA 2081

RESULT 5
ARI36078 2128 bp DNA linear PAT 16-JUN-2001
LOCUS ARI36078
DEFINITION Sequence 1 from patent US 6136580.
ACCESSION ARI36078
VERSION ARI36078.1 GI:14476750
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2128)
AUTHORS Fukuda,M. and Yeh,J.
TITLE beta-1,6-N-acetylglucosaminyltransferase that forms core 2, core 4 and 1 branches
JOURNAL Patent: US 6136580-A 1 24-OCR-2000;
FEATURES
1. 2128
location/Qualifiers
source
BASE COUNT 569 a 477 c 526 g 556 t
ORIGIN

Query Match 85.8%; Score 1990.8; DB 6; Length 2128;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2004; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 244 agatataaagagagcctgaactgttcccttgagacatctatgaatgtcgaagaataacc 303
Db 102 AGATTTAAAGAGAGCCTGAAGCTGTTCTTGACATCTTATGATGTGCAAGAAATATCC 161
Qy 304 tttagagaggttagaagatcagggagacatgtgtgttcacatctgtctgccaaggaacacg 363
Db 162 TTTTGAGAGGTAGAAAGATCAGGGGACATGGTTGTTCACATTTGCTGCCAGGAAACCG 221

Qy 364 ccaagcttcaacttggaaaagaatacagcctgtgtgaagaagatcatccctaaagcaagagag 423
Db 222 CCAGTCTTCACTTGGAAACAGATACGCTTGTGAAGAGATCAATCCCAAGCAGAGAG 281
Qy 424 aagcaactaaagagatgtgtccctccacacttccctgtgtcaggttccacactgtcc 483
Db 282 AAGCTACTAAAGAGATGTGTCTCTCTCAACTTCCCTGTGTCTGCTCCACCTGTCTCC 341
Qy 484 catctgtgaagatgtgttaatgaaagaagactgtgcagctgtgacatctgtgtgtgtgt 543
Db 342 CATCTGTGAGAGATGTTCAATGAAAGAGACTCTCCAGCTGCAATCTACTTGTGGGCTCG 401
Qy 544 ggtgtgtatagt 603
Db 402 GGTGTGTATGT 461
Qy 604 tctgaccacttgggtctgt 663
Db 462 TCTGACCACTTGT 521
Qy 664 tataatctcctgaagaactccagcaagaaggtctatcaactgttcaaggtgtcaaccgag 723
Db 522 TATAATTTCTGAAACTTCCAGCAAAAGAGTCTATCAACTGTTCAGGGGTCAACCGAGG 581
Qy 724 gaccaagaagagagtgcttcaagacttctgaataacccctgaagagtgcaagaagaagag 783
Db 582 GACCAAGAGGCAAGTCTTCAAGGCTATTCTGATTAACCTTGAGAGCTTAAGAGAGAGG 641
Qy 784 ccttcaacagaacaccactactcctccctcaacagaagactgtgaagacttcaagagctgaa 843
Db 642 CTTTCAACAGACACCCTACTCTCTCCACACAGAGACTGTGACACTTCAAGGCTGAA 701
Qy 844 aggaagttcatacagttcccaactgtgagcaagaagaggtgtgtgtgtgtgtgtgtgtgt 903
Db 702 AGGAAGTTCATACAGTCTCCACTGAGCAAAAGAGGAGTGTCCATTTGCAATCTCT 761
Qy 904 atgtgtatctatgaagaagattgaagaacttgaagaagctgtgtgtgtgtgtgtgtgtgt 963
Db 762 ATGTGATTCATGAGAAATTTGAAAGCTTTGAAAGGCTACTGCGAGCTGTGTATCCCT 821
Qy 964 cagaacataactgtgtcatalgt 1023
Db 822 CAGAACATATATCTGTCTCATGTGATGAGAGTCCCAAGAACTTCAAGAGCGGCTG 881
Qy 1024 aaagcaattatctgt 1083
Db 882 AAAGCAATTAATTTCTTGTCTCCAAATGTCTTCAATAGCCAGTAAACCTGTGTGTGT 941
Qy 1084 tatgtctcctgtgtcagaggtgtgaagctgaacctgaactcatgtgaagaagctgtgtcagag 1143
Db 942 TATGCTCTCTGTGTCCAGGCTGCAACTGACCTCAACTGCAATGAGAGACTGTGTCCAGAG 1001
Qy 1144 tcaatgtcgtgtgaataactcctgtgaatacatalgtgtgaagagacttccctataaagaaca 1203
Db 1002 TCAGTGGCGGTGAATTAATCTCTGTAATACATGTGTGAGAGGAGCTTCTTATAAGAGCAAT 1061
Qy 1204 gcaagagatgtgtccaggtgtcctaagaatgtgtgaatgtgtgaagaatgaatgtgtgtgt 1263
Db 1062 GCAGAGATGTCCAGGCTCTCAAGATGTTGAATGGAGAGAAATGATGAGAGTCAAGAGTA 1121
Qy 1264 cctccaaagcaaaagaacccgtgtgaataatacacttgaagtagtgaagagacata 1323
Db 1122 CCTCTTAAGCAAAAGAAACCCGCTGGAATATATCTTTGAGAGTGTGAGAGACATTA 1181
Qy 1324 caactaaccacaagaagaagatcctcccttataatlttaactatgttttaaggaat 1383
Db 1182 CACCTTAACCAACAAGAAAGAGATCTCCCTTTAATTTAATATGTTTACAGGAGAT 1241
Qy 1384 ggtatcatgt 1443
Db 1242 GCGTACATTTGTGCTTCCCGAGATTTGTGCCAAACTGTTTGAAGAACCTTAATATCCAA 1301

QY 1444 caactgataatgagtgtaaaagacattatagccagatgaacacctctggtccacctt 1503
 |||||||
 Db 1302 CAATGATTAATGGGTAAGAAAGACACTTATAGCCAGATGAACACTCTGGGCACTT 1361
 QY 1504 caagcgtcaagcgttgagatgctgtctgtcccaaccacccaagtgcagatctcagac 1563
 |||||||
 Db 1362 CAGGTCACAGGTCGAGGCTGCTGTGTCCCAACACCCCAAGTACGATCTCAGAC 1421
 QY 1564 atgactctatctgcacagcctggtcctcagtggaaggtcatgagagagacatcgataaggt 1623
 |||||||
 Db 1422 ATGACTTCTATTGCGCAGGCTGCTGCACTGCGAGGCTCATAGGAGACATCGATAAGGCT 1481
 QY 1624 gctccctatgctccctgctctcttgaaatcccaacagcggtatctgctttatggtctggg 1683
 |||||||
 Db 1482 GCTCCTTATGCTCCCTCTCTGGAATCCACAGGCGGCTATCTGCTTATAGGGGCTGGG 1541
 QY 1684 gacttgatgagatgctcctcaaacatcacctgtgtgccaacaagtgtgacccaagta 1743
 |||||||
 Db 1542 GACTTGATTTGGATGCTTCAAAACCATCCTGTTGGCCAAAGTTTGAACCCAAAGGTA 1601
 QY 1744 gatgataatgctctcctcagtgcttgaaagaataccatgattataaagccatctatggact 1803
 |||||||
 Db 1602 GATGATATATGCTCTTCACTGCTTGAAGAAATACCTAGTTATAGGCCATCTATAGGACT 1661
 QY 1804 gaactctgagacacacataagagcgttgcctacactgtggtgcaagagcatgtacaaat 1863
 |||||||
 Db 1662 GAACCTTGAGACACACATATAGAGCGTGTGCTACTGTGGGGCAAGACATGATCAAAACAT 1721
 QY 1864 gctcagaactgtctggagacagtggtgtgggagacagcgcttgcgaactcgtgtgactcc 1923
 |||||||
 Db 1722 GCTCAGAACTTGCTGGGACAGTGTGGGTGGAGACAGGCGCTTGGCAATTGCTGGCATCC 1781
 QY 1924 tttagaataagagggctcctatagattgtggtgaatagatcttgccttgcaaatg 1983
 |||||||
 Db 1782 TTTAGGATTAAGAGGGCGTGCATTTAGATTGTGGGTAAGTAACTCTTTGGCCTTGAATTTG 1841
 QY 1984 ctgctgtggtgaatgctgtctgtctcaccacccctaaacccctagtagttccctccactact 2043
 |||||||
 Db 1842 CTGCTCGGGGAATGCGTGTGCTCTCTCAACCCCTAAGCTTAGATGTTCCCTCCACTACT 1901
 QY 2044 ttcccaataagtgagaaatggaactgctgtgatagagagagtgagagagagatgtggtc 2103
 |||||||
 Db 1902 TTCTCACTTAAGTGAATGAGAACTGCTGTGATAGGAGAGTGAAGAGGATATGTGCT 1961
 QY 2104 agagcacttgatcagttgaatgacccctgctgtgactcttccactctgctgagcgtcgt 2163
 |||||||
 Db 1962 AGAGCACTTGATTTCACTTGATGATCCGTGCTGATGCTTTTCCATTTCTGGAGCTGCGCT 2021
 QY 2164 tctcataaatccagagttgtgtagcgtgtgagagagaaacttgatgagaaagaaactcc 2223
 |||||||
 Db 2022 TCCATAAATAATCCAGGTTTGTGTAACGTGGAGAGAACTTGATGAGAAAGAGAAC--TTCC 2079
 QY 2224 ttctgtactgtaactaaataaata 2251
 |||||||
 Db 2080 TTCTGTACTGTTAATAATAATAATA 2107
 |||||||

RESULT 6
 AF102542 2128 bp mRNA linear PRI 31-JAN-1999
 LOCUS AF102542 Homo sapiens beta-1, 6-N-acetylglucosaminyltransferase mRNA,
 DEFINITION complete cds.
 ACCESSION AF102542
 VERSION AF102542.1 GI:4204683
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2128)
 AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.
 TITLE Molecular cloning and expression of a novel beta-1,
 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I

JOURNAL Chem. 274 (5), 3215-3221 (1999)
 MEDLINE 99115671
 REFERENCE 2 (bases 1 to 2128)
 AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.
 TITLE Direct Submission
 JOURNAL Submitted (28-OCT-1998) Glycobiology Program, The Burnham
 Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037, USA
 FEATURES
 source
 1. 2128
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="brain"
 /dev_stage="fetus"
 354.
 /note="glycosyltransferase; core 2 Gnt-M"
 /codon_start=1
 /product="beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="AADI0824.1"
 /db_xref="GI:4204684"
 /translation="MYQMRRLQOLHYLMALAGCYMLIATVALKLSFRLKCDSDHGLIES
 RESQSYCHNIIYNFLPKPKRSINCSGYTRGQDAVLOALINLEEVKKRPETDTH
 YLSIRDCHEFRKAERKEIOEPLSKEVERPIAYSMVTHKEINLEBLAAVAPQITV
 CVHYDEKSPETREKAVKATISCPNPNFASKLVRYVYASMSRQADINCMEDLOSIV
 PMKVFLLNTGCTGFPIKSNAEVQALKMILGRNMSSEVPPKHKETRMKYHFEVVDTL
 HLINRKKDPNPNLMTGONAYIVASRDEVOVHLKPKRSQDLIEWVKDYSPDEHLMA
 TLORARMGSGVSNHPRKYDISMTSISARLVKQGHGDDIKGAPVAPCSGIHQRAICV
 YGADLNWMLQNHHLANKFPRVDNALQCELYRYKATVGTGL"
 BASF COUNT 569 a 477 c 526 g 556 t
 ORIGIN
 Query Match 85.8%; Score 1990.8; DB 9; Length 2128;
 . Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2004; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

|||||
Db 41735 GATTTCAGTTGATGATCTGCTGCTGATGCTTTTCCATTTCTGTGAGCTGCCGCTTCTTAATAA 41794
Qy 2173 ttccaggttgtagtcgcygagagagaactttgatgaaagagaacctcccttcctgtagt 2232
Db 41795 TTCAGGTTTGGTACGTCGGAGAGAACTTGATGAAAGAACCTTCCCTTCGTGACT 41854
Qy 2233 gtaacttaaaataatagctccctgattcaagatataaccttacttcttgctagat 2292
Db 41855 GTTAACCTTAATAATAAGCTCCCTGATTCAAAAGTAATTAACCTTCTTTCCTAGTAAT 41914
Qy 2293 gccagaataataataatcctaacaga 2319
Db 41915 GCCAGAAATAATAATAATAACAGA 41941

RESULT 8
AX087948 1317 bp DNA linear PAT 17-MAR-2001
LOCUS AX087948 Sequence 14 from Patent WO0114535.
DEFINITION AX087948
ACCESSION AX087948
VERSION AX087948.1 GI:13396926
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1317)
AUTHORS Schwientek,T. and Clausen,H.
TITLE Udp-n-acetylglycosamine: galactose--g(b)1,3-n
-acetylgalactosamine--g(a)-r/(glcnac to galnac)-g(b)1,6-n
Patent: WO 0114535-A 14 01-MAR-2001;
Schwientek, Tilo (DK); Clausen, Henrik (DK)
JOURNAL Location/Qualifiers
FEATURES
source 1..1317
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1317
/note="The gene of C2GnT2"
CDS
/codon_start=1
/protein_id="CAC34676.1"
/db_xref="GI:13396927"
/translation="MVQMKRLCQLHYLMALGCMYMLATVALKLSFRLKDSDDLGLSES
RESQSOYCRNLVNFILIPAKRSINSGVMDGDAEYVATLLNLNLYKKREPTDTH
YLSITRDCENFKARKRPIOPPLSKNEVEFPFLAYSMVHEKLEENBRLLRAYAQNIT
CVHDEKSPETFEKAVKAIISCFPNVFLASKLVNVYVASKWSRVOADLMCMEDLLQSSV
PMKYFLNCTGDFPIKSNAMVQALKMLGNRNSMESVPPKHETRMKYHREVVYEDTL
HLTNKKKDDPPYNLTFGTGNAYIVASRDFVQHLKNPKSOQLIEMVKDYSPPDEHLMA
TLORARMMGSPVNPMPKYDIDMTSIALRVKMGHEGDIDKGAAPAPCPSIHORATCY
YGAGDLMMMLQNHLLANKRPKYDDNALQCLEERYLKAITYGTEL"
BASE COUNT 359 a 307 c 322 g 329 t
ORIGIN

Query Match 56.8%; Score 1317; DB 6; Length 1317;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 181 AAACCTCCAGCAAGAGAGGTCTATCAACTGTTCCAGGGGTCACCCGAGGGAGCCCAAGAGGCA 240
Qy 736 gttcttaaggtctatctgataacttgtagtgcaagaagaacgagagcttccagac 795
Db 241 GTGCTTCAGGCTATTTCTGATTAACCTGGAGGTCAGAAAGAACGAGGCTTTCACAGAC 300
Qy 796 acccaactcctccctaccagagacttgagcaactcaaggtcgaaggaagttcata 855
Db 301 ACCCACTACCTCTCCCTCACGAGACACTGTGACACTTCAGAGCTGAAGGAAGTTCAAT 360
Qy 856 cagttcccaactgagcaagaagagtgaggttccctattgcatctatgtagtcaat 915
Db 361 CAGTTCACCACTGAGCAAGAGAGGTGGAGTCCCTATGTGACTATCTATGCTGATTCAT 420
Qy 916 gagaagattgaaacttgaagaggtctgtagctgtagtgcctcccgagacataac 975
Db 421 GAGAAAGATTGAACCTTTGAAAGGCTACTGCGACGCTGTGTATGCCCTCAGAACATATAC 480
Qy 976 tgtgtccatgtagagagaagtcaccagaacttccaagaagcggtcaagaacatatt 1035
Db 481 TGTGTCCATGTGATGAGAGAGTCCCGAGAACTTTCAAAGAGCGGTCAAGCAATTAAT 540
Qy 1036 tcttgcttcccaaatgcttcaatagccagtaagctggttcggtgttatagctccctcg 1095
Db 541 TCTTCTTCCCAAAATGCTTTCATAGCCAGTAAGCTGTGCGGTGATTATGCTCTGCG 600
Qy 1096 tccaaggtgcaagctgacactgaactgcatggaagagcttcccaagctcagtgccgag 1155
Db 601 TCCAGGCTGCAAGCTGACCTCAACTGATGAGAAAGCTTGCTCCAGAGCTCAAGTGCCTGG 660
Qy 1156 aaatacttccatgatacatgtagagcggaacttccatataaagaagaatgcaagatggtc 1215
Db 661 AATATCTTCCATATACATAGTGGAGCGACTTTCCTATTAAGCAATGACATGAGATGCTC 720
Qy 1216 caggtctcaagaatgttgaaatgagagaaatgaacatgagatgacttccatgaac 1275
Db 721 CAGGCTTCACAAATGTTGAATGGAGAGATATACATGAGATGAGAGTACCTCTTAAGCAC 780
Qy 1276 aaagaaccgcctggaataatcactttaggtgagtgagaagaacattacacctaaccac 1335
Db 781 AAGAGAACCCGCTGGAAATATACCTTGAAGTAGAGACACATTACACTTAACCAAC 840
Qy 1336 aagaagaagatccctcccttataatcattcaatcattgattacaggggaatgcatattg 1395
Db 841 AAGAGAGAGATCCCTCCCTTATATTAATTAATTAATGTTACAGGGAATCCGTAFTGCG 900
Qy 1396 gcttcccgagatttcgtccaacatglttgaagaacccttaacccaacactgattga 1455
Db 901 CCTTCCCGAGATTGTGTCACATGTTTGAAGAACCCCTAAATCCCAACACTGATTGANA 960
Qy 1456 tgggttaaaagacacttatagcccaatgaacacctctgggccaaccttcaagtgcaagg 1515
Db 961 TGGGTAAAGACACTTATAGCCCAATGAAACCTCTGGGCCACCTTAGGCTGACGCG 1020
Qy 1516 tgaatgacctggtctglttcccaacacaccccaagtgacacatcagacatgacttcat 1575
Db 1021 TGGATGCCCTGGCTGTCTGTTCACCAACCCAGTAGAGATGTCAGACTGACTTCTAAT 1080
Qy 1576 gccaggtctgtagaagtgaggaagtcatagaggaagacatgataaagggtgcttcatgct 1635
Db 1081 GCCAGGCTGTCAAATGGCAGGTCATGAGGAGACATGCAATGAAGGTCCTCTTAATGCT 1140
Qy 1636 ccctgctcggataccacaggggtctatcgtttatggggtcgggagcttgaattgg 1695
Db 1141 CCTCTCTGGAATCCACAGGGGCTATCTGCTTTAGGGGCTGGGACTTGAATTGG 1200
Qy 1696 atgcttcaaaacatcacactgttggccaacaaggttgaccacaaggtatgatatagct 1755
Db 1201 ATGCTTCAAAACATCACTGTTGGCCAACAGTTTGAACCCAAAGTATGATATATGCT 1260
Qy 1756 cttaagtgttagaagaatcctacgttataaggaacatctatggagctgaacttga 1812
Db 1261 CTTCACTGCTTGAAGAAATACCTACCTATATAGGCCCATCTATAGGAGACTGAACCTTGA 1317

```

RESULT      9
AF038650
LOCUS       AF038650                1317 bp    mRNA    linear    PRI 25-MAR-1999
DEFINITION  Homo sapiens core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase
            (C2/4GnT) mRNA, complete cds.
ACCESSION   AF038650
VERSION     AF038650.1   GI:4511880
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 1317)
AUTHORS     Bennett, E.P., Hollingsworth, M.A. and Clausen, H.
            Control of O-glycan branch formation. Molecular cloning of human
            cDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase
            forming core 2 and core 4
            J. Biol. Chem. 274 (8), 4504-4512 (1999)
            99143102
JOURNAL     MEDLINE
REFERENCE   2 (bases 1 to 1317)
AUTHORS     Schwiensek, T. and Clausen, H.
            Direct Submission
            Submitted (28-NOV-1997) Dental School, Faculty of Health sciences,
            Copenhagen University, Noerre Alle 20, Copenhagen 2200, Denmark
FEATURES
    source
        1..1317
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..1317
            /gene="C2/4GnT"
            1..1317
            /gene="C2/4GnT"
            /note="core 2/4-GnT"
            /codon_start=1
            /product="core 2/core 4
            beta-1,6-N-acetylglucosaminyltransferase"
            /protein_id="AAD21525.1"
            /db_xref="GI:4511881"
            /translation="MYOMKRLQOLHYIMALGCMYLAATVALKILSPRLKODSHLIGLES
            RESQSOYCRNIIILNPLKLPKRKRSINGSGVTRGQEOVLQILNNLEVKKREFPTDTH
            YLSLRDCEHFKERKFTDPLSKSEVEEPITASVTHEITENFERLRLRVAPONTY
            CVHVDKSPETFEKAVKAIISCPNVEIASKLRYVYISWSRYQADLNCMEDLLQSSV
            PWYFLNTGCTDPEIKSNEMVOALKMLNRSMESEVPRKHETKMYFEVARDTL
            HLTNRKKDPPVNLMTFTGNAYIVASRDFOVHLKPKRSOQLIEWKDYSPDEHLMA
            TLQARMMGSSVNHPRKDYDISDMTIARLYKMGHGDIDKGAIPVAPCSIHQRATCV
            YGAGDINMMLQNHHLANKFDPKVDNALQCLEEYLRKATYGTET."
BASE COUNT   359 a      307 c      322 g      329 t
ORIGIN
Query Match      56.8%; Score 1317; DB 9; Length 1317;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 736 gtgtctcaagctatctctgaataaccttgagggtcaagaagaagcgaccttcaacagac 795
    |||||
DB 241 GTGCTTCAGGCTATTCTGAAATTAACCTGAGGCTCAAGAAGAAGCGACCTTTCACAGAC 300
QY 796 acccaactacctccctccaccagaagactgtgagcaacttcaaggtcgaaggaagttcata 855
    |||||
DB 301 ACCCAGTACCTCTCCCTCACAGAGACTGTGACACTTCAAGCTGAAAGAGATTCATA 360
QY 856 cagttccacttgagcaagaagaaggtgaggttccctatgcatactatggtgattcat 915
    |||||
DB 361 CAGTTCCACTGAGCAAGAAGAGAGTGAGATTCCCTATTGCATATCTATGGATTGATCAT 420
QY 916 gagaagaattgaanaacttgaagaagctactgcgagctgtgatagtccctcagaacataac 975
    |||||
DB 421 GAGAGAGATTGAAAACTTTGAAAGGCTACTCGAGCTGTGATGCTCCCTCAGAACATATTAC 480
QY 976 tgtgtcatatgtgatggaagagtcctcccaagaacttcaagaagcggtcgaagaattatt 1035
    |||||
DB 481 TGTGTCCATGTGTGATGAGAGAGTCCCCAGAAACTTTCAAAGAGCGGTCAAGCAATTATT 540
QY 1036 tcttgcttcccaatgtcttcatagccagtaagctgtgtcgggtgttatagtccctcgtg 1095
    |||||
DB 541 TCTTGTCTTCCCAATGCTCTTCATATAGCCATGAGCTGGTTGCGGTGTTATGCTCTCGG 600
QY 1096 tccaaggttgcaagctgagccctcaactgcagtggaagacttgcacagagctcagatgcgtg 1155
    |||||
DB 601 TCCAGGCTGCAAGCTGACCTCAACTGCATGAGACTTGTCTCCAGACTCAGTCCCTCGTG 660
QY 1156 aaactactctgaatacatatgtgagagacttccctctaaagaagcaatggaagatgtgc 1215
    |||||
DB 661 AAATACTTCTGATTAATGTGGAGCGACTTTCCTATTAAGACCAATGACGAGATGTGCT 720
QY 1216 caagctctcaagatgttgaatgaggaagatagatgagatgagctcaaggtaccctcctaagcac 1275
    |||||
DB 721 CAGGCTCTCAAGATGTTGAATGGAGAGATAGATGAGTCAAGAGTACCTCTCAAGCAC 780
QY 1276 aaagaacacccgctggaataatcactttaggttagtgagagacaatacctaaccacac 1335
    |||||
DB 781 AAAAAGAAACCCGCTGGAAATATCACTTGTGAGTGTGAGAACACATTAACCTCAACCAAC 840
QY 1336 aagaagaagatccctccctctaatatgaactgtgttcaaggaatgtgcatatgtg 1395
    |||||
DB 841 AAGAGAGAGATCTCTCCCTTATTAATTATTAATGTTTACAGGGAATGCTATATTGTG 900
QY 1396 gcttcccgagaattcgtccacaacatglttgaagaaccctaatactccacaacatgattgaa 1455
    |||||
DB 901 GCTTCCCGAGATTTCGTCAACATGTTTGAAGAACCCTTAATGCCAACAACGATTGAA 960
QY 1456 tgggttaaaagaacttataagccagatgaacacctctgggccaacctcaagctgtgcagg 1515
    |||||
DB 961 TGGGTAAAGACACTTATATACCCAGATGAACACCTCTGGGCCACCTTCAGCGTGACGG 1020
QY 1516 tggatgtcctgctcgtgttcccaaccacccaagaagtaagacatacgaagacttatt 1575
    |||||
DB 1021 TGGATGCTCTGGCTCTGTTCCTCAACACCCCAAGTAGACATCTCAGACATGACTTTCATT 1080
QY 1576 gccagagcttgatcaagtgagcagggttcatgaggagagacatcgataaagggtgtcctttagct 1635
    |||||
DB 1081 GCCAGGCTGGTCAAGTGCGCAGGGGTCAATGAGGAGACATCATTAAGGTTGCTCTTATGCT 1140
QY 1636 ccctgtcttggaatccaacagcggtactatcgcttatabgggctvgggactggaattgg 1695
    |||||
DB 1141 CCGTGCTCTGGAATCCACACAGCGGGCTATCTGCTTATGGGCTGGGACTGGAATTTGG 1200
QY 1696 atgcttcaaaaacataccctggttgccaacaagtgtgacccaaggtgaagtataatgct 1755
    |||||
DB 1201 ATGCTTCAAAACCATCCACCTGTTGGCCAAACAAGTTTGAACCAAGAGTAGATTAATGCT 1260
QY 1756 cttaagtgttaagaagatacctaagcttaaggaagccatcatvggaactggaacttga 1812
    |||||
DB 1261 CTTCACTGCTTAGAAGATACCTAATGATTAAGGCCATCTATGGGACTGAACATTGTA 1317

```

```

RESULT 10
LOCUS AF318573 108873 bp DNA linear VRL 18-MAY-2001
DEFINITION Bovine herpesvirus 4 long unique region, complete sequence.
ACCESSION AF318573 AF212111
VERSION AF318573.1 GI:12802528
KEYWORDS
SOURCE
ORGANISM Bovine herpesvirus 4.
Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae.
REFERENCE
AUTHORS 1 (bases 1 to 108873)
Zimmermann,W., Broll,H., Ehlers,B., Buhk,H.-J., Rosenthal,A. and
Goltz,M.
TITLE Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
Identification of an Origin of DNA Replication
JOURNAL J. Virol. 75 (3), 1186-1194 (2001)
PUBMED 11152491
2 (bases 1 to 108873)
REFERENCE
AUTHORS Zimmermann,W., Broll,H., Ehlers,B., Buhk,H.-J., Rosenthal,A. and
Goltz,M.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Project P24, Robert Koch-Institut, Nordufer
20, Berlin D-10555, Germany
FEATURES
source
1..108873
Location/Qualifiers
1..108873
/organism="Bovine herpesvirus 4"
/db_xref="taxon:10385"
misc_feature
1..108873
/note="Long unique region; LUR"
complement(445..633)
/note="ORF B01"
CDS
1..108873
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pb01"
/db_xref="GI:12802529"
/translation="MEGDVSMVAFGRKRTRODLNNSAGNEVCCAVLRKNHNVSAG
CAVRRNSPFHSLGKMP"
614..4480
/note="tegment protein; ORF3 BORP1"
/codon_start=1
/product="v_FGAM-synthase"
/protein_id="AAK07922.1"
/db_xref="GI:12802530"
/translation="MDTSPSIOCFEVSRRAPLSNEGOVIOQFTSDQGLPTSLPYQY
EQVPMVYNGEBOAVRELVPLDILFSSHGELTFPHSSILKTSRRSLTYGSPSA
PETALSRDLVAIILOSWSLAPMSMLTQILKIVSIIISIKCLNMPSPKIHSTLRAL
CPDLSLRPKYKGEIIVGAFHTGAPGVLSPPKGDILGAFGLNAGSGASGFMT
POHGPQITVNONFCETAVIVPDYVWEISITVLSGMDTGLATYGGQLOVARDN
LKLFLVSPNOMFNVFSSQLGIASIVSITSPRSSSELPYTRDIDMAVIMMRC
SLPTOGCFRPVYVSHPLDSVPESGPIAFGAVCFSSRNITDLPETRRKNITVQLG
FTSHEDSGREVRCLGALKYMSFLRPOVKGISAVKVCETMAKLLSVCQIGIK
IYOSALPPLNTGLLPPTASNRQNMNTMKHFINIYAVIVLVAIEATPDNTTIVK
AGCAIHCCQYKVLGRQTPHIIIVKDKGEQNVREKISLKSPSYPLSTINKLOKS
ENMODSSMEMDTPLSLTYLLEILKHAVGCKDFIVKHIDRLSGRAVOOQIGARDI
PIDSYILVSLDLSLAAPDREGENPMATASLDPLVNNNEPVPGLCSAIGEEPSY
TYPRIKGAHVAIVSLINIVASAPRRKIDDTCTNITNIPOTDSHFGILDELMGRER
CSQGLICVFTSCTTSNRGKCLNLSLVRLVITAFPCADVQGLPLDEKDDSI
VWLPSTIEHTFPGVISOQLFRDPSGHIINDPLVYKLLITVTKHINHESIISCHV
GCGGLITACEEMAGASIALIVPODEDPLTSETPGILMVEPVKSTVOXKHE
TSDIIFDVGRTPLSVSNFTVSKYKRIIFRESLNEVEMVREHSTKEOMRTYCEY
QCPEDPDKDLHL/TROPSCPHGPKYKHOVNYLLPGTNPDSLLVLAEEGFRN
LVSFTPKYKITYDTTNVPGICLTGATNEDATLGKAIISMTTKHNSVVLGELKLI
NSPDVSLAIGHTAQQLLEFKENKYNKPSKWTYCKENSGLMESRWLNFEIPENH
AVALQSKGSLPGLPWIGTTLGFAPHESETYMETMSTGMAVATQYAGDISAGPATVP
QNPATYTGISGLCSADGRHLALHDPLSLNNLMQWPHI PKMTPLKYSVPMKRMFLDIH
IMANKVREMOQPPPHHPDPLRNLIKVM"
4608..4811
/note="ORF B02; BORFA2"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pb02"
CDS

```

```

CDS
1..108873
/translation="MARELAIVYQVDPDLSKTSILPTCDRHIDILGAFQNNKRLK
LIDALVSLVQNRTECSCLSELEHLITINYITMOKLETLNENITKEDFAMHLIS
DLCEFHKNVNLCEFGDCVPIPSLSIINDIEIFSRLSNVFCISRNNALEETINR
LGTRNISTPIPLDLYPSISCMCLNETSLIPQGEVNVITMNSINCTHATVAPR
PVQGLFEHLMLNIPYDGTGNKVOLETSDRIQNASIEALIDOHNIPEKYSQLEIS
NLIVYNSGHSNDOSPRLCSQMATLILQETPRMAHROKVAASHLALTPRIHFPQSYR
YPIESLRCGIFYSIMDNIOALKQDCSTFLAKSNYKTIHREKMLYRLNNLAPSH
TKESGDIPNTSIOAKRHTIGETSEDSIDAARKNNYIOLAKYTBELKLTIOCLET
QOKIMCNTULRTAGMNYTKAAALIKHNFMIKPAALPQHNHYTAIINELYESKFI
KNSLYLQISOEHDSITLOEYSITLGVGDPFSPPIILANCLADAKVWPHHK
MLISEVIMSWEPMDIDAENAFNYITGKLVNSVOCAMRYIILANCVASVLYRME
KDLIVTSRPTSPNNLSKRPTEFSGIYITTEKAPLFLFINDNGWIFKDIYALLYHH
LOLTSKKNVL"
11420..14044
/gene="gb"
11420..14044
/gene="gb"
/note="ORF 8"
/codon_start=1
/product="glycoprotein B"
/protein_id="AAK07927.1"
/db_xref="GI:12802535"
/translation="MYKTIIFALIKVGSFNQTHSTTSPSSSTSTSTSTSK
PSNTRTNSNSLAASPONTSKSPSTNOGSTRPIPTVDDTAAKNRYKRYNCASS
SGELFRDLDDOTCPDKOKHVEGILVLYLKNKIVPIYFKVAKIKIKATSVYIRGMSO
AAVTNRDISRAIPYNEISMIDRTYHGSAAVYINILNTYIDRDSNSVPLQPAVA
GLTENIRYFQPLIVAEVPGFIVRTVTEENVDVDMARSVEPYTHFTTALGDTI
EISPFCHNNSCCTGNSRDTAVWIEENHQTVDYBRRGHPDKRILFLKDEYTTIS

```

KMAEDREBARA
 LEETVYKTEPPRAIOTI
 HNSFSPYHVAEVA
 AETLNSOEELNGNEI
 LNCANSTINLEETVYK
 KAKRKHOTIRGEVYK
 TITNGCEPLIWOAMK
 PLISEHTNTITI
 ERNNTPGKSRKRSV
 DTPKPGAKLSTPAQOY
 AYDHLTSMNHIL
 EELTKWICRPO
 KKDNLAMWELSKI
 INPVSVAALYEGKVA
 AKAMGADAMVSC
 IINWDAVSJHIMR
 DPVCSYSELVYEV
 NSATPGOAGTNEIL
 THTHETCEPTADH
 EYFVPMKHTYD
 FKDYEVYTMOTN
 ISTIDTFLNLTIDN
 IDICEKIVELYSE
 NERKMAKSLDET
 MERS
 EYNYTOKLASL
 RSDNDTIDNLRN
 DLYVDSLSEMA
 DLDIGICVYNT
 BSGIATYVES
 IYGVGVSEFTIN
 IGDTITLLILV
 VVAYTSRRIN
 NNEAIPKIMY
 PIDASBOE
 EN
 NIOTPGGEIKIL
 LGMOLOSSEHK
 SEESASHKPG
 LEPOLLDBD
 GJLNRRTPL
 IEDPSGNDST
 EHTOKRYV

gene	14184. .17201
	/gene="DPOL"
CDS	14184. .17201

```

/gene="DPOL"
/note="ORF 9"
/codon_start=1
/product="DNA polymerase"
/protein_id="AAK07928.1"
/db_xref="GI:12802536"

```

/transaltion="MEFNPIYSKRTKDKRISVDDDPKRNPKRVLVPCFKTPGPP
 GVIISTITDDPVLPHQDKREYPIETSGDHSORSYSAKLPETAPIDPKRKFHHYDIEE
 TYTDPDRDIPICOFODIPINGVYIKLILGTODASVAVNFOOMYFVYLVGVN
 LSYVLOOTLNGVNRKOTCFKSTIOBKRLIKLEYDAPILVYIKLISAPETLNOIVGN
 FSCGEEVESNNASTREPIINOKRSTFEWCSNHPHISORSDTDLFEPNGCIGDLO
 THEEDOOMPYTIMSEDIIECIGEGGFCFATCDEDEVLVIOCIIMRTDYESPQOIMLS
 LGTQPIANTIDVPEPFSBELDMYFAFTLLRPNDPIEDVGVNIAANDPYIDIRATROY
 NPIPIHFKEVSISSGFEEYNOMNNGAGMRBSFKISGLVADIMVOYCKDLSIOXY
 KINELAKKCMQOKEDYATYKETPIPSFGSGEAGKESVILVADILIKYMHYH
 EITSLAKKAKIDARVULDDQOIRVFSCLTILAKKREHSITILIPBANKRGYQVIGAT
 PLRGYNNPIILVYDASLPSILIOAHNCLISTIMODONMLHHLHPDYETPHLSTVG
 ITHVAKHRTKSLSTILUAMLEKARTIKRELANDONEMGMPKLTILKQALIKVTSNAY
 GTFVAGHKLIPCIPIAETVITLQGRIMLKSAFEMVETPEKLSIOVSVPVCDPDSV
 RIVAYDDSDTELECGYPMOASVLANPADLITPTTKALEFDRILKEAETFOCLIMLKR
 KNYIDILSDSLKRYMGVDLIRKTAQSFVOTTSKELDLVLIDPDKVKAQAOYIQAOPARA
 KYYSGLPVGKRWKVIDLVNNSYALATOKAVESLITESTLSRFEGEKTKTTTPEHLV
 YKRSISRNELFOIHDIRIPVYFKGDGOCGSVDESLAEDEPTYSOANKIPLSVETIYEDKLIL
 HGVANLLOLCIFEGNNSMTVEILYNFVNIPYSEPT"
 17297. , 18577

```

/note="ORF 10: BORFb1: conserved in other
gamma-herpesviruses"
/codon_start=1
/product="unknown"
/protein_id="AAK07929.1"
/db_xref="GI:12802537"
/translation="MTSPILVERSEIEGMQVNICNINILMTNPKDIPVOGGYGLVKKV
FLPLPLNLTNNIVSYSEGLVQVSEFELEDVACQSTAMVQCRFGIPETPLPIPIAHTEFN
MPLFVETQNTPIINSLSLAIIPMKPTCTGCSYSTAPETIPPOHSLAHFV
TGOGHTLDSGAIKCNHKTVMYVFERGSPMOFKEHEHMAGRDLODACVKAQVSD
IKINIMVEVNGLYKAVSVNITNPKHPHPIITLKFTVMVAPDLVAVSYSTPES
GCKVYVITDSEKILKIPGETNLKIKYATVCGNSEKAVMTITSTNNPWTIEPTLS
LMPVPLQVTKIKPTNMIIITIKKDLIAACVAYNSLEEDROPAPSPSYVFNQDILTIV
ESMAVAVILGENTVYSRCHLNINKSENTFSPMDTP"
complement(18848..13945)
/note="ORF B04: short ORF of immediate early transcript 1
RNA"
/codon_start=1

```

Query Match	Similarity	Score	DB	Length
Best Local	80.9%	Pred. No. 4.3e-310	14	108873
Matches 1236	Conservative	0	Mismatches 25	Indels 26
				Gaps 4
OY	441	tgctccctccacccctccctgctgctcggcgtctccacacgtctccccaatctctgacatgct	500	
Db	107264	tgctcctcctccacccctgctgctcggcgtccccaacgtctcctgattccccaatcccaataatggc	107323	
OY	501	tcaatgaaagaga---ctctgcacgtgcatctactgtggtgctcgtggctgctataatgct	557	
Db	107324	tggcgtagaagaagaagactttggccgggagacatcatctgtagggccctggcctgctatatgct	107383	
OY	558	gctgacacatgctggtctcgaacattccttcaaggttgtaagtgctcctgacacattgg	617	
Db	107384	gctggccgctgtgttcttgagagcttttccctcagagttttaaagtgatgagatctcttgga	107443	

QY	618	ttcggatcccaagggaattccaagccagtaactgtgaaataacttctataatttccttga	677
Db	107444	ttcggatcccaagggaattccaagccagtaactgtgaaataacttctataatttccttga	107503
QY	678	acttcacgaagaagttctacaactgttcaagggttcaaccgaaggggaaccaagagcagt	737
Db	107504	acttcacgaagaagttctacaactgttcaagggttcaaccgaaggggaaccaagagcagt	107553
QY	738	gcttcagatcttctgaaataaccttgaaggttcaagaagacgagcctttcaacagac	797
Db	107564	gcttcagatcttctgaaataaccttgaaggttcaagaagacgagcctttcaacagac	107623
QY	798	ccactaaccttccctccaccagaagactgtgacacttcaaggtctgaaaggaagtctaca	857
Db	107624	ccactaaccttccctccaccagaagactgtgacacttcaaggtctgaaaggaagtctaca	107683
QY	858	gttcacacttgaagaagaagaggttgaagttccctattgtacatactctatgttata	917
Db	107684	gttcacacttgaagaagaagaggttgaagttccctattgtacatactctatgttata	107743
QY	918	gaagaattgaaaaacttgaagagctacttgcgagctgtgtatgtccctcagaacatactg	977
Db	107744	gaagaattgaaaaacttgaagagctacttgcgagctgtgtatgtccctcagaacatactg	107803
QY	978	tctccatgttgatagagaaggtccccaagaacttcaaaagagcggttcaagaactatttc	1037
Db	107804	tctccatgttgatagagaaggtccccaagaacttcaaaagagcggttcaagaactatttc	107863
QY	1038	tgtcttcccaaatgtcttcatatagccaagtaagctgtgttcggtgttatagttccttcgttc	1097
Db	107864	tgtcttcccaaatgtcttcatatagccaagtaagctgtgttcggtgttatagttccttcgttc	107923
QY	1098	cagggltgcaagcttgaacctcaactgtcatatgaaagacttgcctcagaagctcagttgcgttgaa	1157
Db	107924	cagggltgcaagcttgaacctcaactgtcatatgaaagacttgcctcagaagctcagttgcgttgaa	107983
QY	1158	atacttctgaaatactgttgggaaggaacttctcataaagaagcaatgcagaatgtgtcca	1217
Db	107984	atacttctgaaatactgttgggaaggaacttctcataaagaagcaatgcagaatgtgtcca	108043
QY	1218	agctctcaagaatgttgaatgggaagaatagcatatgagttcagaagttacctcctaagacaa	1277
Db	108044	agctctcaagaatgttgaatgggaagaatagcatatgagttcagaagttacctcctaagacaa	108103
QY	1278	agaagaaccgcttgaagaatactcaacttgaagttatgagagacacatacctaaccatacaaa	1337
Db	108104	agaagaaccgcttgaagaatactcaacttgaagttatgagagacacatacctaaccatacaaa	108160
QY	1338	gaagaagaatcctccctcctaataattcaactagtcttaacaggaatgtctacattgtgtgc	1397
Db	108161	gaagaagaatcctccctcctaataattcaactagtcttaacaggaatgtctacattgtgtgc	108220
QY	1398	tcccgagaattctgcacacatgttttgaagaacccttaaacctccacaacacttgaatg	1457
Db	108221	tcccgagaattctgcacacatgttttgaagaacccttaaacctccacaacacttgaatg	108280
QY	1458	ggtaaaagaacacttataagcccagaatgaaacactcttgggccaaccttcaaggttgcaggtg	1517
Db	108281	ggtaaaagaacacttataagcccagaatgaaacactcttgggccaaccttcaaggttgcaggtg	108340
QY	1518	gatgccttggtctgttctccacaaccccgaagtaacagacatctcaagatgaactctatgtc	1577
Db	108341	gatgccttggtctgttctccacaaccccgaagtaacagacatctcaagatgaactctatgtc	108400
QY	1578	cagagctgttcaagatgtgcaggggtcatatggggagacatcgataaagggtgtctctatgtctc	1637
Db	108401	cagagctgttcaagatgtgcaggggtcatatggggagacatcgataaagggtgtctctatgtctc	108460
QY	1638	ctgcctctgaatcccaacgagcggtcatctgtgttatactggagcttgggaacttgaattgat	1697
Db	108461	ctgcctctgaatcccaacgagcggtcatctgtgttatactggagcttgggaacttgaattgat	108520
QY	1698	gttcaaaaacataccctgttgcaccaaaagtgttaccaccaaggttgaatgataatgtctt	1757

Db	Accession	Source	Length	Strain	Year	Location
Db 108521	TCTCCAGAACCATCACTCTTTGGCAAACAGTTTGACCCGAGGGGTGATGATTAACGTCT	GenBank	108580			
Oy 1758	tcagtcgtctagaagaatacatctacgtltataaagccatcatcatgagactgagacttgaagac	GenBank	1817			
Db 108581	GCAGTGGCTTAAAGAGTACCTACGTCATTAAGGCCATCATATGGACTGACATCTTTGACCTGC	GenBank	108640			
Oy 1818	actatgagagcgttctactctgtggtggcaagaagcatgtacaacatgctcagaacttgc	GenBank	1877			
Db 108641	ACTGT-----GGGCAGAGACTGTGGCAACATGCCACGAGGGGTACT	GenBank	108681			
Oy 1878	gggaacaggt-gggttgggaagaccaggttgcgaattcggtgcacacctttagaagaag	GenBank	1936			
Db 108662	GGTACTGTGTGGGGGGGGGAGGAGGCGCTGTGGAAATCATGATCCATCCCGAGGGGTACAG	GenBank	108741			
Oy 1937	ggtctcatatgattgtgtgtaagtag	GenBank	1963			
Db 108742	GGCTGCTTTGATGGGCTCAGCATGTTC	GenBank	108768			
RESULT 11						
AF231105	2017 bp DNA linear VRL 26-MAY-2000					
LOCUS	Bovine herpesvirus 4 beta-1,6-N-acetylglucosaminyltransferase					
DEFINITION	(BOREF3-4) gene, complete cds.					
ACCESSION	AF231105					
VERSION	AF231105.1					
KEYWORDS	GI:8096688					
SOURCE	Bovine herpesvirus 4.					
ORGANISM	Bovine herpesvirus 4.					
REFERENCE	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae.					
AUTHORS	1 (bases 1 to 2017) Vanderplasseten, A., Markine-Gorlaynoff, N., Lomonte, P., Suzuki, M., Hiraoka, N., Yeh, J. C., Bureau, F., Willems, L., Thiry, E., Fukuda, M. and Pastoret, P. P.					
TITLE	A multipotent beta-1,6-N-acetylglucosaminyl-transferase is encoded by bovine herpesvirus type 4					
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (11), 5756-5761 (2000)					
MEDLINE	2 (bases 1 to 2017)					
REFERENCE	Vanderplasseten, A., Markine-Gorlaynoff, N., Lomonte, P., Suzuki, M., Hiraoka, N., Yeh, J. C., Bureau, F., Willems, L., Thiry, E., Fukuda, M. and Pastoret, P. P.					
AUTHORS	Direct Submission					
TITLE	Submitted (03-FEB-2000) Immunology-Vaccinology, University of Liege, Faculty of Veterinary Medicine B43b1s, Liege B-4000, Belgium					
JOURNAL	Location/Qualifiers					
FEATURES	1..2017					
Source	/organism="Bovine herpesvirus 4"					
	/strain="V. test"					
	/db_xref="taxon:10385"					
	227..1549					
	/gene="BOREF3-4"					
	227..1549					
	/gene="BOREF3-4"					
	/function="forms core 2, core 4, and I branches"					
	/note="last ORF at the right end of the L-DNA; smaller to Homo sapiens core 2					
	beta-1,6-N-acetylglucosaminyltransferase, mucin type; pBOREF3-4"					
	/codon_start=1					
	/product="beta-1,6-N-acetylglucosaminyltransferase"					
	/protein_id="AAF22001.1"					
	/db_xref="GI:8096689"					
	/translation="MKKAGKKKLCPCGHHLMALGYLWLAIVSLRLSLRFKCDVSDID					
	LESDFQSOHCRLMLYSLKLPAPRSINCGLIRGDEAVQALDMLVKKRPPPLT					
	DYVLTATRCDECRKAORKEIOPPLSKSEIDFLPAYEMVYHEKTEENRILRAYAO					
	NYICVHWLVKSPERFEKAVKAIISCPENVMASGLPVVVASRVOADLNCMEDLIQ					
	SSVWKVLCNTGCTDFPIKTNAEMVLAIKMLKGNMESFSPSSKRNKKRYREYTD					
	TLPTSIRKPPNDLPMFTGNATFTASRAFLVCHVLDNPKSQILLVENWKDYSDEHL					
	WATLQRRPMWGSVPSPKRYHIDSMATIALVWKQYHGVSGAPVAPDSGIHRAI					
	CITGAGGLVYLQHNHLLANKPFPBVDVNVYLCIEELIRKAIYGTGL"					

[illegible]

Db 1135 CTCTCGAGCCCTTGTTCACATGCTTAAAGAACCTTAAATCCCAATPACTGGTTGAATG 1194
 Qy 1458 ggtaaagacactatagaccagatgacacaccttgggacaccttaagcgtgaagcgtg 1517
 Db 1195 GGTAAAGACACATATAGCCCGACGACACACCTCTGGGCCACCTTACAGCTGCTCCGTG 1254
 Qy 1518 gatgcctgctcgttcccaacaaccccaagttacagacatctagacatgacttattgc 1577
 Db 1255 GATGCCCTGGCTGTCTGCTTCTGACCCCAAAAGTATCATCTCAGACATGAGCTGCATCGC 1314
 Qy 1578 caagctggtcaagtggaaggtgcatgaggagacatgataaggtgtctcttattgc 1637
 Db 1315 CAGGCTGTCAAGTGGCAAGTACACAGAGGAGATGTACAGCATGGGGCGCTTATGACAC 1374
 Qy 1638 ctgtctcgaatccaccagcagcgtatctcgtttatcgggctgggactgtaattgat 1697
 Db 1375 CTGCTCTGGAAATCATGGAGGCGCATCTGCATTACGGGGCGGGACCTGTACTGGAT 1434
 Qy 1698 gcttcaaacacatcacctgttggccaacaagttgaccacaaggtatgataatgctct 1757
 Db 1435 TCTCCAAAGACCATCCTCTTGGCAACAGTTTGACCCGAGGGTGATGATACGTCCT 1494
 Qy 1758 tcaagtgttgaagaataactaagttataagggcactctatgggactgaactttgagacac 1817
 Db 1495 GCACTGCTTGAAGAGTACTACATCATAGGCCATCTATGGGACTGAACCTTTGA----- 1549
 Qy 1818 actatgagagcgttgcctacgttgggccaagagcatgtacaacatgctcagaacttct 1877
 Db 1550 -----GCTGCACAGGTGGGCCAAGGACTGTGCCAACATGCCAGGGCGTACT 1555
 Qy 1878 gggacagatgt-ggglyggagacagggccttgcacatcgttgcacatcttgaataag 1936
 Db 1596 GGTACTGTGTGGGTGGGGGCGAGGGCTGTGGAATCCATGACATCCCGCAGGGTCAAG 1655
 Qy 1937 ggcctgcatatgattgtgggttaagtag 1963
 Db 1656 GGTGCTTCATGGGTGGGTGACAGATGTTG 1682
 RESULT 12
 BHAAREGF 4179 bp DNA linear VRL 06-JUL-1995
 LOCUS Bovine herpesvirus type 4 DNA for nonconserved region F (DMS99 like strain).
 DEFINITION 246386
 ACCESSION 246386
 VERSION 246386.1 GI:562278
 KEYWORDS
 SOURCE Bovine herpesvirus 4.
 ORGANISM Bovine herpesvirus 4.
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 Gammaherpesvirinae.
 REFERENCE 1 (Bases 1 to 4179)
 AUTHORS Lomonte, P.
 TITLE Direct Submission
 JOURNAL Submitted (20-OCT-1994) Patrick P.L. Lomonte, Virology, University
 of Liege, Faculty of Veterinary Medicine, Bd de Colonster, Sart
 Tilman, Liege, 4000, BELGIUM
 REFERENCE 2 (Bases 1 to 4179)
 AUTHORS Lomonte, P., Bublot, M., van Santen, V., Keil, G.M., Pastoret, P.P. and
 Thilly, E.
 TITLE Analysis of bovine herpesvirus 4 genomic regions located outside
 the conserved gammaherpesvirus gene blocks
 JOURNAL J. Gen. Virol. 76 (Pt 7), 1835-1841 (1995)
 MEDLINE 97201604
 FEATURES
 source Location/Qualifiers
 1. 4179
 /organism="Bovine herpesvirus 4"
 /db_xref="taxon:10385"
 /clone="Ecorti H (Bublot et al., 1990) tot"
 BASE COUNT 1192 a 927 c 951 g 1109 t

Query Match 41.4%; Score 960; DB 14; Length 4179;
 Best Local Similarity 80.9%; Pred. No. 4.1e-294;
 Matches 1180; Conservative 0; Mismatches 245; Indels 33; Gaps 4;
 Qy 507 gaagaagactctgcacagctgattactatgtgtgctctggtgtctatatgtctgacac 566
 Db 2649 GAAGAAGCTTTGGCCGGGAGCATCTGTGTGGCCCTTGCTGCTATATAGCTGCTGGCGCT 2708
 Qy 567 tgtgctctgaactcttcttcccaagttgaagtgtgactcgtgacacactggtctgagtc 626
 Db 2709 TGTTCCTCTGAGATCTTCCCTCAGGTTTAATGATGATGATGATCTTGGATCTGAGATC 2768
 Qy 627 caaggaaactcnaagccagtaactgtgaatatcttgtataattccctgaactccagc 686
 Db 2769 CAGGAGCTTCACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 2828
 Qy 687 aaagaagctatcaactcgttccagaggttcacccgagggagaccagaagagcaatgtctaacgc 746
 Db 2829 AAAAGATCCATCAACTGTTCTGGATCAGCGAGGGGACCAAGCAAGTGTGTCCAGGC 2888
 Qy 747 tattcgaataactgtgaggttcaagaagaagcgaagccttccacagacccactact 806
 Db 2889 CTCTCTAGACACCTGSAAGTGAAGAAAAGCGGTGCTCTCTCACTGGACATTTACT 2948
 Qy 807 ctccctacacagagactgtgagcaacttcaaggtcgaagaaagttcalatagttccact 866
 Db 2949 CAACATPACCAAGAGACTGTGAGCGCTTCAAGGCCCAAGAGAGTTCATACAGTTCACAT 3008
 Qy 867 gaagcaagaagaaggtgaggttccctatgtcatctatctatgtgtatgataagaagatga 926
 Db 3009 GAGTAAAGAGAGGTACACTTCCCATTTGCTTCTCATGTGTGTATGAGAAAGATGA 3068
 Qy 927 aaacttgaagagctactgtgagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 986
 Db 3069 AAACCTTGAACGGCTGCTGCGAGCGGTGTATGCCCCCTCAACATATATCTGTGCCAGCT 3128
 Qy 987 ggaatgagaagtlcccccagaaacttcaagaagcggatcaaaatlaatttctgtctccc 1046
 Db 3129 GGAATGAGAGTCCCGAGAGACTTTCAAAAGAGGGCGTCAAGGCCATPATTCTGCTTCC 3188
 Qy 1047 aaatgtcttatagccagtaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1106
 Db 3189 CAAGTCTTCAATGCGCGTAAGTGTGTCGGGTGTATATGCTCCCTGTCAGAGATGA 3248
 Qy 1107 agctgacctcaactgcataagagactgtctccagaagctcagtcggtggaataactctc 1166
 Db 3249 AG-----CTGTATGAGAAAGCTTGCTCCAGAGCTCACTAGTGCATGGAAGTACTACT 3298
 Qy 1167 gaatatactgtggaacggaacttccctataaagaagaatgacagatgttccagctctcaa 1226
 Db 3299 GAATACATGCGGGGACAGACTTCCCATTAAGAACCAATGCGAGATGCTCTGCGCCCTCAA 3358
 Qy 1227 gatgttgaatggagagatgacatgagtgatgagatgacatgacatgacatgacatgacatg 1286
 Db 3359 GATGTGTAAGGGTAAAGACGTATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3418
 Qy 1287 ctggaataatcaacttgaagtagtgagagacacatcaactcaacacacacagaagaaga 1346
 Db 3419 CTGGAATATACCGCTATAGAG---GTGACAGACACTGTATACCTACACGAAAGATGAAGA 3475
 Qy 1347 tctctccctataatataatctatgatttacaaggaatgagatgagcttgcctccgaga 1406
 Db 3476 CCTCCGCCCTGATATATATACCATGTTCAACAGGAATAGCTATATTTGTGCGCTCGAGC 3535
 Qy 1407 ttctgtccacaatgttttgaagaacctataatcccaacactgattgaaggttaagaaga 1466
 Db 3536 CTTTGTCTCAACATGTCTTAAACAACTTAATCCCAAGAGCTGTTGAAGTGGTCAAGA 3595
 Qy 1467 caactataagccagaatgaacaccttgggcaaccttcaagcgtgacaggtgaggtgctcgtg 1526
 Db 3596 CACCTATAGCCCGACAGCAACCTCTGGGCCACCTTCAAGCGTGTCTCTGTGATGCTGG 3655


```

Oy 1527 ctctgtcccaaccaccccaagtaacacatcagacatgacttatttccagcagctggt 1586
Db 3656 cctcttctcttaccacaccccaatcacttctcagacatgacttgcacatgccacgctgct 3715
Oy 1587 caagtgcagggttcattgaggagacatcgataagggtctcttattgtctccctgtctg 1646
Db 3716 CAAAGTGGCAGTACCAACGAGGAGATGTCAGCATGGGGGGCGCTTATGACACCTGCTGG 3775
Oy 1647 aatccaccagcggcgtatctgcttattgaggcttggaggaactggaatgattgattcaca 1706
Db 3776 AATCCATGGAGGGCCATCTGTATTTACGGGCGCGGGCCCTGTACTGTGATCTCCAGAA 3835
Oy 1707 ccatcactgttgcgcaacaagtattgacccaaggtatagataatgtctcttcacgtgctt 1766
Db 3836 CCATACCTCTTGCGCAACAATTGACCCGAGGGGTGATGATACGTCTCTCAGAGTCTT 3895
Oy 1767 agaaataactactgattataaggcaactatagggactgaactttagacacacatagaga 1826
Db 3896 AGAAGAGTACCTACCTCATTAAGCCATCTATGCGACTGTAACCTTGCAGCTGCACCTGT- 3951
Oy 1827 gctgttaccctgtgggagcaagacatctacaaacatgctgaactgtctggagcagt 1886
Db 3952 -----GGGCCAGGAGACTGTGCCAACAATGCCAGGGGTACTGTGTACTGTG 3996
Oy 1887 t-ggggtggagagaccagggcttgcacattcgtgcacatcccttaggataagaggctgcat 1945
Db 3997 TCGGGTGGGGGAGCAGAGGGGTGTGATTCATGTCATGCCATCCCGGAGGTCAAGGGGTGCTT 4056
Oy 1946 tagattgtgggttaagtag 1963
Db 4057 GATGGGGTCAAGATGTGTG 4074

```

RESULT 13

BC018297

LOCUS BC018297 1841 bp mRNA linear ROD 06-DEC-2001

DEFINITION Mus musculus, RIKEN cDNA 2010013H22 gene, clone MGC:18826

IMAGE:4208500, mRNA, complete cds.

ACCESSION

BC018297

VERSION BC018297.1 GI:17390702

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1841)

REFERENCE

AUTHORS

JOURNAL

TITLE

DIRECT SUBMISSION

Submitted (03-DEC-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

REMARK

COMMENT

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalob@bcm.tmc.edu

Villalob, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia,

A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,

Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov> Series: IRAK Plate: 24 Row: p Column: 13 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES

source

Location/Qualifiers

1..1841

/organism="Mus musculus"

/db_xref="LocustID:72077"

/db_xref="taxon:10090"

/clone="MGC:18826 IMAGE:4208500"

/tissue_type="Colon, normal. 5 month old male mouse."

/clone_id="NCI CGAP_C024"

/lab_host="DH10B"

/note="Vector: pCMV-Sport6"

192..1388

/codon_start=1

/product="RIKEN cDNA 2010013H22 gene"

/protein_id="AAH18297.1"

/db_xref="GI:17390703"

/translation="MDLSEERQSOYCDLLKTLKLPKASSINCSYIRBQKAVTO

ALUNLETKRQOULETADYDLMTADCEHFTKRFIOVPLSKESAPFIASVYAS

KIENRELLRAVYTPONVYCVHMDQKSEPKOAVRIVSCFPNVFIASKSVVYAS

MSRVOADLNCEMEDLOSPPMKYLLNTGDEPIKTNAEMKALKLKGMSMESEVP

PPHKSRRKHYEVDTLHMTSKRTPPNLTMFTGNAYVVASDFIEHVSNSKAR

OLIEVVKPTVSPDEHMTATLORASMPGSDPLHRRFDLSDMRAIRLTKVYDHEGDI

NGAPYTSQSGIHQRAVCYGSGLHWIIONHHLANKFDPKVDNVLOCLEBYLHKA

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

IYTEL"

Query Match 39.8%; Score 923.8; DB 10; Length 1841;
Best Local Similarity 73.1%; Pred. No. 1.2e-282;
Matches 1307; Conservative 0; Mismatches 447; Indels 35; Gaps 8;

BASE COUNT 496 a 448 c 469 g 428 t

ORIGIN

```

Oy 481 tcccatctctgacgagtggttcacatggaagagacgtccagcgtgacattctgtggt 540
Db 60 tccctgcctctgtaagatgacttcttcgacagagcttctgacacatgacctgtgacac 119
Oy 541 ctgggtcgtatatactgctggtccactgtgctctgaaacttcttcaggttaagtg 600
Db 120 CTGGGTGCTACATGCTACTGACGACATCCCTTGCCTGAAACTGTCCCTCAGACTGAAGTGT 179
Oy 601 gactctgaccacttggtgtgtgagtcgaagagatctcgaagcaagctagtagaatc 660
Db 180 GACTTCGATGCCATGATGTGACCTCTGAGGAATTTCAAGCCAGTACAGGGATCTC 239
Oy 661 ttgataaatcttcctaactccagaagaaggtctataactgttcaagggtccaccga 720
Db 240 CTGTACAAACCCCTGAGCTGACGACGACGACGACGACGACGACGACGACGACGACGAC 299
Oy 721 ggggaccagaagagcagtgcttcaagctatctgaaaccttggaaggtcaagaagaagcga 780
Db 300 GGGGAGCAGAAAGCGGTGACCCAGGCTCTGCAATTAACCTGGAATTAAGAAAGAGCAG 359
Oy 781 gaggcttccagaacacccactacactctccacagaagaggttgaagacttcaaggt 840
Db 360 CAGCTCTTTCACAGAGCCGACCTACTTACGATGACGACGACGACGACGACGACGACGAC 419
Oy 841 gaaagaagttatcaagttccactgaagcaagaagaggttgaagttccacttgcacac 900
Db 420 AAGAGGAATTTATACAGTCCCACTGACGACGACGACGACGACGACGACGACGACGACGAC 479
Oy 901 tctatgtgattcatcagaagattgaaacttgaagacttgcagagctgtgtatgtgc 960
Db 480 TCCATGTGTGTCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 539
Oy 961 cctcagaacatactactgtgtcattgtgataagaggtcccccagaacttcaagaagagcg 1020
Db 540 CCTCAGAAATGTACTGTGTCCACATGATGATGATGATGATGATGATGATGATGATGATGAT 599
Oy 1021 gtcaagaacaattattcttcttcccaaatgcttcacagcagtaagctgtgtggtg 1080
Db 600 GTCAAGGCGCATGTGTATGCTTCCCAATGTCTTCAATGCTGATGATGATGATGATGATGAT 659
Oy 1081 gttagtctctgttccagaggtgcaagctgacactcaactgattgaagactgtgtccag 1140

```

D	660	GTCTATGTTCTTCTGTCACAGGATGACGGCTGACCTTAACGTGATGGAAGACTTGCTTCAG	719
O	1141	agctcagtgccgtfygaaatactctctcgaatacattgtggagcygacttctctataaagc	1200
D	720	AGCCCCGGCCATGGAATACCTCTCGAACAACCTGTGGAGACAGACTTTCCTCAAAACC	779
O	1201	aatcagaagatggtccagagctcccaagatgtgaaatggtggaagaaagacgtgaagcagag	1260
D	780	AATGCTGAGATGGTCAGAGCCCTCAGACTATTGAAAGGGCAGACAGATGTGAGTCAAG	839
O	1261	gtaccctcgaagcaaaagaaccgcgtgnaaalacatttgaggtagtgagacaca	1320
D	840	GTACCACCTCCACATATAAAATCCCGCTGGAAATATCACTATGAG---GTGACACACACA	896
O	1321	ttacaactaaccaagaagaagatgcctccccctataatttaactatgtttcagag	1380
D	897	TTGCACATGACACACAGAGAGAGACGCCGCCACCTAATTAACCTATACATGTTCACGTGG	956
O	1361	aatgctacatgtgagcttcccgagatlttcgtccaacatglttltgaagaaacctlaacc	1440
D	957	AATGCTACATAGTGGCTTCTCGAGACTTCAATGAAACGTGTGATGATACCTCAAAAGCC	1016
O	1441	caacaactgattgaaatggtgtaaaagaaacttataagcccaatgaaacactctggtccac	1500
D	1017	CGGCAACTGATCGATGGTGTAAAAAGACACTATAGTCCGATAGACACCTTTGGGCCACC	1078
O	1501	cttcagcgtgacggttgaatgagctctgctctcccaacacaccccaagtaagacatcca	1560
D	1077	CTCCAGGGTGCTGTGGATGCTGGATGATGATCCCTTGCAATCGAAATTTGACTGTGCA	1136
O	1561	gacatgacttctatctgcacagctgtgcaagtggcaaggtcatatgaggagaatcgataag	1620
D	1137	GACATTAGAGGCCATTTGGAGAACTAACCAATGGTACACCATGAGGAGACATTTAGAAC	1196
O	1621	ggtgtcccttatgtgccctctctggaatccaaacagcgggtcatctgctgttatggact	1680
D	1197	GGGGACCTTACACGCTTGTCTGACGAAATCCACAGGGGCTGTCTGTATTATGGTCA	1256
O	1661	ggggaaacttgaaatgtagtcttcaaaaacatcacctgttggccaagaagtlttaaccaaag	1740
D	1257	GGGGACCTCGACTGTGATCACTTCAAGAACCAATCACTTTGGCCAAAGATTTCACCAAG	1316
O	1741	gtagaataatagtctctgaagttcttaagaagaatacctaacttataagaagcaataagg	1800
D	1317	GTGATGTATATGTCTTCAAGTGTTTAGAAAGATATTATTCGTACAAACCATCTATGAG	1378
O	1801	actgaactttagaacaactatgaaagcgtgtctactcgttggggcaagaagcatgtacaa	1860
D	1377	ACTGAACATATAG--ACACTGTAAAGACTATTGTACTCTGTGAGCAAGAAATG-----	1428
O	1861	catgctagaactctgctgggaacagttgtgggtggagaaacaaaggtcttgcattgttgca	1920
D	1429	-----CCAGGCACTACCTGGGGGAGGTGGGTGTGAATGTGGGGCTCTAAGCTCATATG	1483
O	1921	tcctttagaataaagggcgtcatatagattgttgggtaagtatgattcttgcgtgcaa	1980
D	1484	TCCCTTAGAGAGAGGGGCGTATTGAGAGAGGGGTAATAGTAACCTGCTCCTTGGA	1543
O	1961	tttgtctgctgvgtgaatgctgtgttctctcaacccttaaccctagtagtcttccacta	2040
D	1544	--GGTGTGCGGTGGAAGGTGGCTATTTCCACCTCCCTCCCA-----CTTCTTTTG	1594
O	2041	acttcttaactaaatgaaatgaaatgaaatgctgtgataagg-----agatgaaagagagat	2096
D	1555	GCTTCTCGCAGACCTTGAAACATGAATACTGTTAATTGAGATTAATAGGAGGGGACAT	1654
O	2097	atgtgtagagcaacttgattcaagtgtgaatgctcgtctgttgaacttttcatctcgttgg	2156
D	1655	GTGACACAGGACCTGTGATTTTGGCTGAATACTCTGTCAGATT-----CTCTACAG	1709
O	2157	ctgcgcgttctcaataaattccaagttgttgtagcgttggaggaagaactttagtgaagaaga	2216
D	1710	ACTTTGTTTCAAAATCCTGTGGGGGGTGTGCTTTGGGTGGGGAACCTTATAGTCAAAAGGC	1769

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 05:58:21 ; Search time 375.65 Seconds
(without alignments)
10599.027 Million cell updates/sec

Title: US-09-874-390-1
Perfect score: 2319
Sequence: 1 ataatgggtttctcatt.....ataataaactcaacaga 2319

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802:*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2319	100.0	2319	21	AAAA8623
2	2205	95.1	2236	22	AAH34463
3	2198.6	94.8	2229	21	AAAG9109
4	2023	87.2	2108	21	AAA96570
5	1990.8	85.8	2128	22	AAAG2134
6	1317	56.8	1317	21	AAA96569
7	1221	52.7	1221	21	AAA96575
8	942.2	40.6	997	19	AAV59800
9	486.2	21.0	777	19	AAV59681

10	456	19.7	524	21	AAA96571
11	369.4	15.9	2204	19	AAAX24043
12	369.4	15.9	2204	21	AAAC65469
13	367.8	15.9	2105	19	AAO61559
14	364.6	15.7	5010	19	AAAX24042
15	364.6	15.7	5010	19	AAAC65468
16	355	15.3	2102	18	AAT80112
17	284.8	12.3	300	21	AAA01261
18	250	10.8	267	21	AAA96572
19	199	8.6	997	22	AAAS4282
20	191.8	8.3	1362	22	AAAS0045
21	180.4	7.8	1807	16	AAQ89201
22	180.4	7.8	1807	17	AAT16201
23	180.4	7.8	1807	19	AAV30006
24	180.4	7.8	1807	19	AAV16000
25	162.8	7.0	378	17	AAT16202
26	139.6	6.0	410	22	AAAS4357
27	139.6	6.0	410	22	AAI85732
28	134	5.8	378	19	AAV30005
29	133	5.7	210	21	AAAG98742
30	129.6	5.6	192	22	AAAG2139
31	84	3.6	147	22	AAAG2142
32	69	3.0	69	21	AAAG6574
33	47.2	2.0	387	22	AAAF6613
34	46.4	2.0	781	22	AAH98678
35	43.6	1.9	66	17	AAT16204
36	43	1.9	99	17	AAT16203
37	39.8	1.7	189	22	ABA65512
38	39.8	1.7	189	22	ABA35576
39	39.8	1.7	189	22	AAK14936
40	39.8	1.7	189	22	AAK40669
41	39.8	1.7	189	22	AAI21432
42	39.8	1.7	189	22	AAI46721
43	39.8	1.7	189	22	AAI07127
44	39.8	1.7	434	22	ABA35937
45	39.8	1.7	454	22	ABA23688

ALIGNMENTS

RESULT	ID	AAA48623	standard; CDNA; 2319 BP.	Noncoding region o
1	AAAA8623	standard; CDNA; 2319 BP.	Human core 2GNT DN	Human core 2GNT DN
19-SEP-2000	AAAA8623;	(first entry)	Human heart core 2	Human heart core 2
19-SEP-2000	AAAA8623;	(first entry)	CDNA sequence of h	CDNA sequence of h
Human C2/4GNT CDNA.			Rat DHI CDNA.	Rat DHI CDNA.
Human; C2/4GNT; UDP-N-acetylglucosamine; O-glycan biosynthesis;			Diabetic rat heart	Diabetic rat heart
O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.			Coding sequence fo	Coding sequence fo
			Human colon cancer	Human colon cancer
			Noncoding region o	Noncoding region o
			DNA encoding novel	DNA encoding novel
			Human DNA encoding	Human DNA encoding
			I-branching enzyme	I-branching enzyme
			Beta-1,6-N-acetylgl	Beta-1,6-N-acetylgl
			Full length CDNA s	Full length CDNA s
			Human beta-1,6-N-a	Human beta-1,6-N-a
			C2GNT catalytic do	C2GNT catalytic do
			DNA encoding novel	DNA encoding novel
			Human polynucleoti	Human polynucleoti
			CDNA encoding a re	CDNA encoding a re
			Human colon cancer	Human colon cancer
			Murine beta-1,6-N-	Murine beta-1,6-N-
			Murine beta-1,6-N-	Murine beta-1,6-N-
			Core 2 beta-1,6-N-	Core 2 beta-1,6-N-
			Novel human polyinu	Novel human polyinu
			Human EST-derived	Human EST-derived
			C2GNT C-terminal r	C2GNT C-terminal r
			Human foetal liver	Human foetal liver
			Probe #12042 for g	Probe #12042 for g
			Human brain expres	Human brain expres
			Human bone marrow	Human bone marrow
			Probe #11365 for g	Probe #11365 for g
			Probe #15407 used t	Probe #15407 used t
			Probe #7118 used t	Probe #7118 used t
			Human foetal liver	Human foetal liver
			Probe #2154 for ge	Probe #2154 for ge

PR 04-DEC-1998; 98DK-0001605.
 XX (CLAU/) CLAUSEN H.
 XX Clausen H, Schwientek T;
 XX WPI; 2000-423407/36.
 DR P-PSDB; AAY94492.
 XX
 XX New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
 PT probe for the detection of specified glucosaminyltransferase from
 PT other species and related organisms
 XX
 XX Claim 5; Fig 2; 47pp; English.
 XX
 CC The present sequence encodes human UDP-N-acetylglucosamine:
 CC N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
 CC (C2/4GNT). The protein is the third member of the family of O-glycan
 CC beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
 CC based O-glycans on to oligosaccharides, glycoproteins and
 CC glycosphingolipids. C2/4GNT can therefore be used in the production of
 CC appropriately glycosylated glycoconjugates with particular enzymatic,
 CC immunogenic, or other biological or physical properties. The nucleotide
 CC sequence is useful as a probe for the detection of C2/4GNT from other
 CC species and related organisms and for the recombinant production of
 CC C2/4GNT polypeptide. The nucleotide sequence was identified by analysis
 CC of EST database sequence information. Oligonucleotides derived from EST
 CC clone 17665 of ATCC were used to isolate two full-length C2/4GNT clones
 CC from a human foreskin genomic PL library by 5' RACE PCR. RT-PCR was
 CC performed using Colo205 human cell line mRNA in order to produce cDNA
 CC for expression of C2/4GNT in Sf9. The control of O-glycan core assembly
 CC has been implicated in tumour progression and metastasis.
 CC
 XX
 XX Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other;
 SQ

Query Match 100.0%; Score 2319; DB 21; Length 2319;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attaacgggtttccattatctatcctctgcattactctctgaagagccctc 60
 DB 1 attaacgggtttccattatctatcctctgcattactctctgaagagccctc 60
 QY 61 tctcttaagtcacgggaactgcccctgtctactgtgacactgccccttactcaacatt 120
 DB 61 tctcttaagtcacgggaactgcccctgtctactgtgacactgccccttactcaacatt 120
 QY 121 ttgtctcgggaagccctcgggaattctgctaatacctacatgtagtgctgaaggaaac 180
 DB 121 ttgtctcgggaagccctcgggaattctgctaatacctacatgtagtgctgaaggaaac 180
 QY 181 agatgaagaatctactcaagagcttctcgtcaatgagaagccaaagcgaagcctcg 240
 DB 181 agatgaagaatctactcaagagcttctcgtcaatgagaagccaaagcgaagcctcg 240
 QY 241 caaagatatataaaggagcctggaactgttctcttgacatctataatgaatgcagaaat 300
 DB 241 caaagatatataaaggagcctggaactgttctcttgacatctataatgaatgcagaaat 300
 QY 301 acctttggagaggttagaagatcaggggaatagttgttcaacatttctgcacgagaa 360
 DB 301 acctttggagaggttagaagatcaggggaatagttgttcaacatttctgcacgagaa 360
 QY 361 ccggcacttctacttgaagaacagaaacagccttggagaagatccatcccaagcagga 420
 DB 361 ccggcacttctacttgaagaacagaaacagccttggagaagatccatcccaagcagga 420
 QY 421 gagaagactactaaagattgtctctcctcaacttccctgtgtcgtcccaactgtc 480
 DB 421 gagaagactactaaagattgtctctcctcaacttccctgtgtcgtcccaactgtc 480
 QY 481 tcccatctctgacgatgtgtcaatggaagagactctgcacgtcattactgtggct 540

DB 481 tcccatctctgacgatgtgtcaatggaagagactctgcacgtcattactgtggct 540
 QY 541 ctgggcctcatatgctgcgcgcacgtgtcctctgaacttcttcacagtgaatgtc 600
 DB 541 ctgggcctcatatgctgcgcgcacgtgtcctctgaacttcttcacagtgaatgtc 600
 QY 601 gactctgacacttggtgtcgtgagtcagaggaaatcacaagccagctctgaaatatac 660
 DB 601 gactctgacacttggtgtcgtgagtcagaggaaatcacaagccagctctgaaatatac 660
 QY 661 ttgtataatttctctgaacttccagcaagaaggtctatcaactgttgaaggttaccga 720
 DB 661 ttgtataatttctctgaacttccagcaagaaggtctatcaactgttgaaggttaccga 720
 QY 721 ggggaccaaaggagagtgcttcaagctatctctgaataacctggaggtcagaagaagcga 780
 DB 721 ggggaccaaaggagagtgcttcaagctatctctgaataacctggaggtcagaagaagcga 780
 QY 781 gaggcttcaagaacaccactactctctcctcaccagagactgtgacacttcaaggtc 840
 DB 781 gaggcttcaagaacaccactactctctcctcaccagagactgtgacacttcaaggtc 840
 QY 841 gaaaggaagtcatagagttcccaactgagcaagaagagtgaggttccctatgtcatc 900
 DB 841 gaaaggaagtcatagagttcccaactgagcaagaagagtgaggttccctatgtcatc 900
 QY 901 tctatgtgtttcaatgagaagatgaaacttgaagagctacgtcgaagctgtgtatgac 960
 DB 901 tctatgtgtttcaatgagaagatgaaacttgaagagctacgtcgaagctgtgtatgac 960
 QY 961 cctcagaacatatactgtgtccatgtgtagatgagagtcaccagaacttccaaagagcg 1020
 DB 961 cctcagaacatatactgtgtccatgtgtagatgagagtcaccagaacttccaaagagcg 1020
 QY 1021 gtcaagcaattattcttctgtctcccaatgcttcaatagccagtaagctgtgtcgggt 1080
 DB 1021 gtcaagcaattattcttctgtctcccaatgcttcaatagccagtaagctgtgtcgggt 1080
 QY 1081 gttatgctctcctgtgtccaggtgcaagctgacactcaactgcatggaagctgtctcag 1140
 DB 1081 gttatgctctcctgtgtccaggtgcaagctgacactcaactgcatggaagctgtctcag 1140
 QY 1141 agctcagtgccgttgaaatacttctcgaatacatgtgtagagcggaacttccataaagac 1200
 DB 1141 agctcagtgccgttgaaatacttctcgaatacatgtgtagagcggaacttccataaagac 1200
 QY 1201 aatgcagaagatgtgtccaggtctcgaagtgttgaatgaggaggaatagcatggaagcag 1260
 DB 1201 aatgcagaagatgtgtccaggtctcgaagtgttgaatgaggaggaatagcatggaagcag 1260
 QY 1261 gtaaccttaagcaagaagaaacccgttggaatatcaacttggagtagtgaagagacaca 1320
 DB 1261 gtaaccttaagcaagaagaaacccgttggaatatcaacttggagtagtgaagagacaca 1320
 QY 1321 ttacaccttaaccaagaagaagatcctccctataatthaactatgtttacaggg 1380
 DB 1321 ttacaccttaaccaagaagaagatcctccctataatthaactatgtttacaggg 1380
 QY 1381 aatgcytaacattgtgtgtctccgagatttgcgtcaaatgttttgaagaacccctaaatcc 1440
 DB 1381 aatgcytaacattgtgtgtctccgagatttgcgtcaaatgttttgaagaacccctaaatcc 1440
 QY 1441 caaacaactgttaagtggaagaagacactatagaccagatgaacacccctgggcacc 1500
 DB 1441 caaacaactgttaagtggaagaagacactatagaccagatgaacacccctgggcacc 1500
 QY 1501 cttaagctgacaggttgatgagctgtctgttcccaaccaaccacccaagtagacatctca 1560
 DB 1501 cttaagctgacaggttgatgagctgtctgttcccaaccaaccacccaagtagacatctca 1560
 QY 1561 gacatgactcttatgtccagagctgtgtcaagtgtgcaaggtcatggaaggagacatcgaatg 1620


```

Db 1561 gacatgactctatgccaagctggtlcaagtgccagggctcatgagggagacatgataag 1620
Oy 1621 ggtgctcttatgctcctgctcgtcgtgaatccaccagcgggctatcgtgttaagggt 1680
Db 1621 ggtgctcttatgctcctgctcgtcgtgaatccaccagcgggctatcgtgttaagggt 1680
Oy 1661 ggggactgtaattgagatgcttcaaaacacatcacctgttgcgaacaagttaaccacaag 1740
Db 1661 ggggactgtaattgagatgcttcaaaacacatcacctgttgcgaacaagttaaccacaag 1740
Oy 1741 gtagatgataatgctcttcagtgcttagaagaatacctacgttataagggcaatcaggg 1800
Db 1741 gtagatgataatgctcttcagtgcttagaagaatacctacgttataagggcaatcaggg 1800
Oy 1801 actgactcttagacacactcttagagcggttgctacccctgtgggggaaagaatgatacaaa 1860
Db 1801 actgactcttagacacactcttagagcggttgctacccctgtgggggaaagaatgatacaaa 1860
Oy 1861 catgctcagaactctgtcggacagctgtggtggtggagaccagggcttgcgaatctggca 1920
Db 1861 catgctcagaactctgtcggacagctgtggtggtggagaccagggcttgcgaatctggca 1920
Oy 1921 tcccttaagataagaagggcgctctatagatgtgtggtaagtagatcctttgccttgcaa 1980
Db 1921 tcccttaagataagaagggcgctctatagatgtgtggtaagtagatcctttgccttgcaa 1980
Oy 1981 ttgctgctgggtggaatgctctgtcttctcaacccctaacccctagatgcttccacata 2040
Db 1981 ttgctgctgggtggaatgctctgtcttctcaacccctaacccctagatgcttccacata 2040
Oy 2041 accttctacataagtgagaatgagaactcgtgtgtaagtagaagtgagaagggagatagt 2100
Db 2041 accttctacataagtgagaatgagaactcgtgtgtaagtagaagtgagaagggagatagt 2100
Oy 2101 ggtgagagactgtgatttcaagttgaatgctcgtcgtgtgaactttccattctgtgagctgc 2160
Db 2101 ggtgagagactgtgatttcaagttgaatgctcgtcgtgtgaactttccattctgtgagctgc 2160
Oy 2161 cgttccataataatccaggttctgtagcgttgagagagacttgaatggaagaagacctt 2220
Db 2161 cgttccataataatccaggttctgtagcgttgagagagacttgaatggaagaagacctt 2220
Oy 2221 cccctctgactgttaacttaaaaaataaatagctccctgatccaagaattaccctctactt 2280
Db 2221 cccctctgactgttaacttaaaaaataaatagctccctgatccaagaattaccctctactt 2280
Oy 2281 ttgctcgtatgtagcagaataataataatacttaacaga 2319
Db 2281 ttgctcgtatgtagcagaataataataatacttaacaga 2319

RESULT 2
AAH34463
ID AAH34463 standard; cDNA: 2236 BP.
AC AAH34463:
XX
XX 03-SEP-2001 (first entry)
DT
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
DE
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 15; ss.
OS Homo sapiens.
XX
XX MO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.

```

```

PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Barash SC, Birse CE, Rosen CA:
PI
XX WPI: 2001-235357/24.
DR
XX P-PSDB: AMG75058.
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 1; Page 3190-3191; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAG77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other:
SO

Query Match 95.1%; Score 2205; DB 22; Length 2236;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 64 ctctaaagtcagggagactcctcctgctactgtgacgcctccttaactcaagcattttg 123
Db 2 ctctaaagtcagggagactcctcctgctactgtgacgcctccttaactcaagcattttg 61
Oy 124 ttctgggaagccttggtgattcgtcctaatacctacacgtgagtgctgaagggaacaga 183
Db 62 ttctgggaagccttggtgattcgtcctaatacctacacgtgagtgctgaagggaacaga 121
Oy 184 tgaagaacatgacccctcaaggaagcctcctgctcaatagagaagaccagcctggca 243
Db 122 tgaagaacatgacccctcaaggaagcctcctgctcaatagagaagaccagcctggca 181
Oy 244 agatataaagagagccttgaaactgttcccttggaacatctatagaaatgacgaataacc 303
Db 182 agatataaagagagccttgaaactgttcccttggaacatctatagaaatgacgaataacc 241
Oy 304 ttggaaggttaagaagatcagggagacatgtgttcacatttgcgccaggaacacg 363
Db 242 ttggaaggttaagaagatcagggagacatgtgttcacatttgcgccaggaacacg 301
Oy 364 ccagctcacttggaagaacagaatacagccttggaagaagatcacccttaagcagaagag 423
Db 302 ccagctcacttggaagaacagaatacagccttggaagaagatcacccttaagcagaagag 361
Oy 424 aagctactaaagattgtgtcctcctcacccttccctgtgtcgttcacactgtctcc 483
Db 362 aagctactaaagattgtgtcctcctcacccttccctgtgtcgttcacactgtctcc 421
Oy 484 catctgtgacgatgttcaatggaagagactcgcagcgtgcatctactgttgggtctg 543
Db 422 catctgtgacgatgttcaatggaagagactcgcagcgtgcatctactgttgggtctg 481

```

QY 544 gctgctatatgctgctgacacgtggtctctgaacattctctcaggttgaaagtgtgac 603
|||||
Db 482 ggtcgtatatgctgctgacacgtggtctctgaacattctctcaggttgaaagtgtgac 541
QY 604 tctgacacacttggtctggtggtccaggaattccaaagccagtactgttaagaatactttg 663
|||||
Db 542 tctgacacacttggtctggtggtccaggaattccaaagccagtactgttaagaatactttg 601
QY 664 tataattctctgaacattccagcaaaaggttctataacgttccagggtgaccccgagg 723
|||||
Db 602 tataattctctgaacattccagcaaaaggttctataacgttccagggtgaccccgagg 661
QY 724 gaaccaagagcagtgctcaggtctatctgaataaccttggaagtgcaagaaagcgagag 783
662 gaaccaagagcagtgctcaggtctatctgaataaccttggaagtgcaagaaagcgagag 721
QY 784 ccttccagaacacccctcctcctccctccagaagagctgagccctccagaagcgtaa 843
|||||
Db 722 ccttccagaacacccctcctcctccctccagaagagctgagccctccagaagcgtaa 781
QY 844 aggaagttcatacagttccacactgagcaaaagaggtggaattccctattgcaactct 903
|||||
Db 782 aggaagttcatacagttccacactgagcaaaagaggtggaattccctattgcaactct 841
QY 904 atggttgatcatagaagaattgaaacctttgaaagcctactcgagctgtgtatgcccc 963
842 atggttgatcatagaagaattgaaacctttgaaagcctactcgagctgtgtatgcccc 901
Db 964 cagaacatacagtggtcagctgtgagtagaagtcctccagaacattccagaagcggtc 1023
902 cagaacatacagtggtcagctgtgagtagaagtcctccagaacattccagaagcggtc 961
QY 1024 aaagcaattattctctgctccccaatgltctcatacagtaagcgtgtgtcgggtgtt 1083
962 aaagcaattattctctgctccccaatgltctcatacagtaagcgtgtgtcgggtgtt 1021
QY 1084 tatgctccctggtccagaggtgcaagctgacactgaactgcatggaagcctgtctccaaagc 1143
1022 tatgctccctggtccagaggtgcaagctgacactgaactgcatggaagcctgtctccaaagc 1081
Db 1144 tcaagtgcgtggaatactctctgaatacattggtgagcggaacttctcataaagagaat 1203
1082 tcaagtgcgtggaatactctctgaatacattggtgagcggaacttctcataaagagaat 1141
QY 1204 gcagagatggtccaggtctccaagatgttgaatggaagaaatagcatggaagtgcaagagta 1263
1142 gcagagatggtccaggtctccaagatgttgaatggaagaaatagcatggaagtgcaagagta 1201
QY 1264 ccttccataagacaagaagaaacccgttggaataatcaactttgagtgagagagacacatta 1323
1202 ccttccataagacaagaagaaacccgttggaataatcaactttgagtgagagagacacatta 1261
QY 1324 cacttaacacaagaagaagatcctccctccctataatttaactatglttaccaggggaat 1383
1262 cacttaacacaagaagaagatcctccctccctataatttaactatglttaccaggggaat 1321
Db 1384 gcgtacaattgtgcttcccgagatttcgtccaacatggttttgaagaaccttaateccaa 1443
1322 gcgtacaattgtgcttcccgagatttcgtccaacatggttttgaagaaccttaateccaa 1381
QY 1444 caactgttgaatggtggaagaagacattatagccagataaaccctctggccaacctt 1503
1382 caactgttgaatggtggaagaagacattatagccagataaaccctctggccaacctt 1441
QY 1504 cagcgttgaagtgatgctggtctgcttccccaaccccaagtgacgacatccagac 1563
1442 cagcgttgaagtgatgctggtctgcttccccaaccccaagtgacgacatccagac 1501
QY 1564 atgacttctatltgcagagctgttcaagtggaagtgcatgagggagacatcogaataaggt 1623
1502 atgacttctatltgcagagctgttcaagtggaagtgcatgagggagacatcogaataaggt 1561
QY 1624 gctccattatgctccctgctctgtgaatccaccagcggtctctcgtcttatggggtcggg 1683

Db 1562 gctccattatgctccctgctctgtgaatccaccagcggtctctcgttcttatggggtcggg 1621
QY 1684 gacttgattggaatgcttcaaaacacacactccttctggccaacaagtctgacccaagagta 1743
1622 gacttgattggaatgcttcaaaacacacactccttctggccaacaagtctgacccaagagta 1681
QY 1744 gatgatatagtctctcagtgcttagaagaatactaatgattatagaagccactatgaggaat 1803
1682 gatgatatagtctctcagtgcttagaagaatactaatgattatagaagccactatgaggaat 1741
Db 1804 gaacttgaagacacacataagaagcgttgcatactgttgggcaagagcagatcacaacat 1863
1742 gaacttgaagacacacataagaagcgttgcatactgttgggcaagagcagatcacaacat 1801
QY 1864 gctcagaacttgctggagacagtggtgggtggagacacagccttgcaattcgtgacacc 1923
1802 gctcagaacttgctggagacagtggtgggtggagacacagccttgcaattcgtgacacc 1861
QY 1924 tttagaataagagggcgtcattagatgtggttaagtagatcttctgcttgcacaaatg 1983
1862 tttagaataagagggcgtcattagatgtggttaagtagatcttctgcttgcacaaatg 1921
QY 1984 ctgcttggtggaatgctgcttcttctcaaccccttaaccttaagtctctccactaact 2043
1922 ctgcttggtggaatgctgcttcttctcaaccccttaaccttaagtctctccactaact 1981
QY 2044 ttctcactaaagtgaagaaatggaactgctgctgataagagaggttaagagagataatgtgt 2103
1982 ttctcactaaagtgaagaaatggaactgctgctgataagagaggttaagagagataatgtgt 2041
Db 2104 agagcaattgattcagtgaaatgctgctgctgtaagcttctccattctgtgagctgcgt 2163
2042 agagcaattgattcagtgaaatgctgctgctgtaagcttctccattctgtgagctgcgt 2101
QY 2164 tctcctaattccaaggttggtgagcgtggaaggaacttgatggaagaaagaaacttccc 2223
2102 tctcctaattccaaggttggtgagcgtggaaggaacttgatggaagaaagaaacttccc 2161
Db 2224 ttctgactgttaactaaataaataatagctcctgattcacaagta 2268
2162 ttctgactgttaactaaataaataatagctcctgattcacaagta 2206

RESULT 3
AAC99109
ID AAC99109 standard; cDNA; 2229 BP.
XX
AC AAC99109:
DT 09-MAR-2001 (first entry)
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
XX
XX Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
XX detection; diagnosis; identification; cytosolic; neuroprotective;
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US05989.
PF 12-MAR-1999; 99US-0124270.
PR
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;
PI
XX MPI: 2000-579444/54.
DR P-PSDB: AAB54344.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 1: Page 759-760; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytosolic,
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC proteins and cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;

Query Match 94.8%; Score 2198.6; DB 21; Length 2229;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2198; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 70 gtacagggaactgcccctgctactgtgacctgccccttactcaagcagtttgtcttg 129
DB 1 gtacagggaactgcccctgctactgtgacctgccccttactcaagcagtttgtcttg 60
QY 130 gaagccctggtattctgtcaatacctaatacctgtagtgctgaagggaacagatgaaga 189
DB 61 gaagccctggtattctgtcaatacctaatacctgtagtgctgaagggaacagatgaaga 120
QY 190 acatgacctcaaggagcttccctgtcaataggaagaccagcgtgacgctggcaagaat 249
DB 121 acatgacctcaaggagcttccctgtcaataggaagaccagcgtgacgctggcaagaat 180
QY 250 taaaggagagcctgaaactgttcccttggaacattatgaatgtcagaataacaccttgg 309
DB 181 taaaggagagcctgaaactgttcccttggaacattatgaatgtcagaataacaccttgg 240
QY 310 aaggttaagaagatcaggagacatgtgttcaacattgtctgcacaggaacccgcagatc 369
DB 241 aaggttaagaagatcaggagacatgtgttcaacattgtctgcacaggaacccgcagatc 300
QY 370 ttcacttggaacacagatacagcctgtgaagaagatcacccctaaagcagaagagacta 429
DB 301 ttcacttggaacacagatacagcctgtgaagaagatcacccctaaagcagaagagacta 360
QY 430 ctaaaagattgttccctccctcaacttccctgtgctcgggtctccacattctcccaattc 489
DB 361 ctaaaagattgttccctccctccctccctgtgctcgggtctccacattctcccaattc 420
QY 490 gtagcagatgttcaatagtagagagactctgcagcgtcattactgtgggtctctggctgc 549
DB 421 gtagcagatgttcaatagtagagagactctgcagcgtcattactgtgggtctctggctgc 480

QY 550 tataatgctgtgcccactgtgctctgaaacttcttcaaggtgagtgactctgac 609
DB 481 tataatgctgtgcccactgtgctctgaaacttcttcaaggtgagtgactctgac 540
QY 610 cacttggtctgagtcacagggaatctcaagcagactgtatgaatatctgtataat 669
DB 541 cacttggtctgagtcacagggaatctcaagcagactgtatgaatatctgtataat 600
QY 670 ttccctgaacttccacagaagaggtctatcaactgttccgggttccccgagggagccaa 729
DB 601 ttccctgaacttccacagaagaggtctatcaactgttccgggttccccgagggagccaa 660
QY 730 gaggcagctgtccaaagcttctcgaataacccctgaaagtcagaagagcagagccttc 789
DB 661 gaggcagctgtccaaagcttctcgaataacccctgaaagtcagaagagcagagccttc 720
QY 790 acagacaccactactctccctcaccagagactgtgagcaacttcaaggctgaaagaa 849
DB 721 acagacaccactactctccctcaccagagactgtgagcaacttcaaggctgaaagaa 780
QY 850 ttcatatcaattccacttgcaagaagaggtgaggttccctattgcatatctatgtg 909
DB 781 ttcatatcaattccacttgcaagaagaggtgaggttccctattgcatatctatgtg 840
QY 910 attcaatgagaagatgaaacttgaagagctactgcagctgtgatgcccctcagaac 969
DB 841 attcaatgagaagatgaaacttgaagagctactgcagctgtgatgcccctcagaac 900
QY 970 atatactgtgtccatgtgagtgaagaagtcaccagaacttccaaaggcgtgtcaaga 1029
DB 901 atatactgtgtccatgtgagtgaagaagtcaccagaacttccaaaggcgtgtcaaga 960
QY 1030 atatttctgtctcccaaatgtcttccatagccagtgaaagctgtgttggtttatg 1089
DB 961 atatttctgtctcccaaatgtcttccatagccagtgaaagctgtgttggtttatg 1020
QY 1090 tctgtgtccaggtgtgcaagctgacccctcaactgcatgtgaagactgtgtccagagctag 1149
DB 1021 tctgtgtccaggtgtgcaagctgacccctcaactgcatgtgaagactgtgtccagagctag 1080
QY 1150 ccgttggaataacttccctgaatacatatgttggaacgagacttccctataaaggcaatgacag 1209
DB 1081 ccgttggaataacttccctgaatacatatgttggaacgagacttccctataaaggcaatgacag 1140
QY 1210 atgtgtccaggtctcgaagatgttgaatggggaataagcatggaagtcagatctct 1269
DB 1141 atgtgtccaggtctcgaagatgttgaatggggaataagcatggaagtcagatctct 1200
QY 1270 aagcacaaagaagaccgctggaataatactacttgaaggtgaggaacacattacacta 1329
DB 1201 aagcacaaagaagaccgctggaataatactacttgaaggtgaggaacacattacacta 1260
QY 1330 acccaacaagaagaagatcctcccttataatlaactatgttitaacaggaatgctgac 1389
DB 1261 acccaacaagaagaagatcctcccttataatlaactatgttitaacaggaatgctgac 1320
QY 1390 atgtgtgttcccggaattctgtccaaatgttgaagaacaccttaalcccaacaacg 1449
DB 1321 atgtgtgttcccggaattctgtccaaatgttgaagaacaccttaalcccaacaacg 1380
QY 1450 attgaaatggtaaaagacactataagccagatggaacacctctgggcaacaccttcaggt 1509
DB 1381 attgaaatggtaaaagacactataagccagatggaacacctctgggcaacaccttcaggt 1440
QY 1510 gaacggtgagatgctgtgttcccaaccaccccaagtaacacatcctcagacatgact 1569
DB 1441 gaacggtgagatgctgtgttcccaaccaccccaagtaacacatcctcagacatgact 1500
QY 1570 tctattgcaagctgtgtcgaagtgcagggatcatgaaaggagacatcgataaggtgtcct 1629
DB 1501 tctattgcaagctgtgtcgaagtgcagggatcatgaaaggagacatcgataaggtgtcct 1560

QY 1630 tatgtccctgtctctgtgaatccaccagcggtctactctgtttagtgggctcggaacttg 1689
|||
DB 1561 tatgtccctgtctctgtgaatccaccagcggtctactctgtttagtgggctcggaacttg 1620
|||
QY 1690 aattggtgtcttcaaacatcacctctgttgccaacaagttagcccaagaagtatgat 1749
|||
DB 1621 aattggtgtcttcaaacatcacctctgttgccaacaagttagcccaagaagtatgat 1680
|||
QY 1750 aatgctcttaagtgcttagaagaatacttaagcttataagggcactcttgaggaactt 1809
|||
DB 1681 aatgctcttaagtgcttagaagaatacttaagcttataagggcactcttgaggaactt 1740
|||
QY 1810 tgagacacactatgagagcgctgttactctgtgggccaagagcatgtacaacatgtctag 1869
|||
DB 1741 tgagacacactatgagagcgctgttactctgtgggccaagagcatgtacaacatgtctag 1800
|||
QY 1870 aactgtctgggacagtgctgtgggtgggagacagggcttgcaatctgtggcactcttag 1929
|||
DB 1801 aactgtctgggacagtgctgtgggtgggagacagggcttgcaatctgtggcactcttag 1860
|||
QY 1930 ataagagggcgctctattagattgtgggtaagtagatcttggcttgcaaatgtctgct 1989
|||
DB 1861 ataagagggcgctctattagattgtgggtaagtagatcttggcttgcaaatgtctgct 1920
|||
QY 1990 ggggtgaatgtcgtctgtctctccacccttaaccctagtagtctccactaacttctca 2049
|||
DB 1921 ggggtgaatgtcgtctgtctctccacccttaaccctagtagtctccactaacttctca 1980
|||
QY 2050 ctaagtgaagatgagaaactgtgtgtagagagtagaagagtagaagagagatagtgtagagca 2109
|||
DB 1981 ctaagtgaagatgagaaactgtgtgtagagagtagaagagtagaagagagatagtgtagagca 2040
|||
QY 2110 ctgatttcagtttgaatgacgctcgtgtagacttccattctgtgtgagctgcgtctctaa 2169
|||
DB 2041 ctgatttcagtttgaatgacgctcgtgtagacttccattctgtgtgagctgcgtctctaa 2100
|||
QY 2170 taattccaggtttgtgtagcgtgtagaggaactttgatggaagaagaacctccctctgt 2229
|||
DB 2101 taattccaggtttgtgtagcgtgtagaggaactttgatggaagaagaacctccctctgt 2160
|||
QY 2230 actgttaactaaataataatagctccgtgattcaaaagta 2268
|||
DB 2161 actgttaactaaataataatagctccgtgattcaaaagta 2199
|||

RESULT 4
AAA96570
ID AAA96570 standard; DNA: 2108 BP.
XX
AC AAA96570;
XX
DT 08-FEB-2001 (first entry)
XX
DE A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.
XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLcNAc-T;
KM cancer; cardiovascular disorder; inflammatory disorder; asthma;
KM rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KM septic shock; adult respiratory distress syndrome; ARDS; cancer;
KM platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KM clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KM diverticulitis; ulcerative colitis; ss.
XX
XX Homo sapiens.
OS
XX
PN CA2296936-A1.
XX
PD 03-AUG-2000.
XX
PF 03-FEB-2000; 2000CA-2296936.
XX
PR 03-FEB-1999; 99US-0118674.
XX

PA (GLYC-) GLYCDESIGN INC.
XX
PI Korczak B, Lew A;
XX
XX WPI; 2000-594746/57.
DR
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
XX
XX Claim 4; Page 51-52; 66pp; English.
XX
XX
XX The present sequence encodes a partial human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GLcNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
XX Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;
SQ

Query Match 87.2%; Score 2023; DB 21; Length 2108;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 atattaaagagagcctgaaactgttctctggacattatgatgtcagaataactct 305
|||
DB 18 atattaaagagagcctgaaactgttctctggacattatgatgtcagaataactct 77
|||
QY 306 ttgagaggttagaagatcagagggacatggtgttcacatttgcgcagcaggaacggcc 365
|||
DB 78 ttgagaggttagaagatcagagggacatggtgttcacatttgcgcagcaggaacggcc 137
|||
QY 366 agtctcaacttgaaacagaatccacgccttctgaaagatcatccctcaagcagagagaa 425
|||
DB 138 agtctcaacttgaaacagaatccacgccttctgaaagatcatccctcaagcagagagaa 197
|||
QY 426 gctactaaagatgtgtctcctccacactccctgtgtcgtgtccacactgtctcca 485
|||
DB 198 gctactaaagatgtgtctcctccacactccctgtgtcgtgtccacactgtctcca 257
|||
QY 486 ttctgtacagatgttcaatgaaagaaactctgccaactgcatattgtgggtctggg 545
|||
DB 258 ttctgtacagatgttcaatgaaagaaactctgccaactgcatattgtgggtctggg 317
|||
QY 546 ctgtataatcgtgtgcgaactgtgtgtcgtgaaacttctcttcaggttgaagtgtgact 605
|||
DB 318 ctgtataatcgtgtgcgaactgtgtgtcgtgaaacttctcttcaggttgaagtgtgact 377
|||
QY 606 tgaccacttggcttggagtcacaggaatcccaagcagactgttagaataatctgtga 665
|||
DB 378 tgaccacttggcttggagtcacaggaatcccaagcagactgttagaataatctgtga 437
|||
QY 666 taatttccaaactccacgcaaaagaggtctataactgttcaggggtcacccagaggga 725
|||
DB 438 taatttccaaactccacgcaaaagaggtctataactgttcaggggtcacccagaggga 497
|||
QY 726 ccaagagcgagtgcttcaaggtatctgtaataacttggaagtcagaagaagcagagggc 785
|||
DB 498 ccaagagcgagtgcttcaaggtatctgtaataacttggaagtcagaagaagcagagggc 557
|||

QY	786	tttaagagacacccactactctctcccttaacccagagactgtgagcacttcaagctcgtgaag	845
Dp	558	tttcaacagacacccactactctctccctacccagagactgtgagcacttcaagctcgtgaag	617
QY	846	gaagttcatacaattcccaactgtgacaaagaaggtggagttccctattgtcatactcat	905
Dp	618	gaagttcatacaattcccaactgtgacaaagaaggtggagttccctattgtcatactcat	677
QY	906	ggtgatctcagtgaaagatttgaanaaacttgaagaagctactgagactgtgtatgtccctca	965
Dp	678	ggtgatctcagtgaaagatttgaanaaacttgaagaagctactgagactgtgtatgtccctca	737
QY	966	gaacatactatgtgtccatgtgtgaatgaagaatccccaagaacttcaagaagcggtca	1025
Dp	738	gaacatactatgtgtccatgtgtgaatgaagaatccccaagaacttcaagaagcggtca	797
QY	1026	agcaattattctgtcttcccacaaagtctctcaagcagtaagctgtgttggtgttta	1085
Dp	798	agcaattattctgtcttcccacaaagtctctcaagcagtaagctgtgttggtgttta	857
QY	1086	tgacctctgtgtccaggggtgcaagctgaacttcaacatgtgacatgagaagactgtgtccagagctc	1145
Dp	858	tgacctctgtgtccaggggtgcaagctgaacttcaacatgtgacatgagaagactgtgtccagagctc	917
QY	1146	agtgccgtgtgaaatactctccatgatacatatgtgtgagcagacttccctataagaagcaatg	1205
Dp	918	agtgccgtgtgaaatactctccatgatacatatgtgtgagcagacttccctataagaagcaatg	977
QY	1206	aggaatgtgtccaaagctctcaagatgtgtgaatggaggaatagcatggaggtcaagaggtacc	1265
Dp	978	aggaatgtgtccaaagctctcaagatgtgtgaatggaggaatagcatggaggtcaagaggtacc	1037
QY	1266	tcctaaagcacaaagaagaaacccgctgtgagaataatcaacttggaggtagtgagaacacattca	1325
Dp	1038	tcctaaagcacaaagaagaaacccgctgtgagaataatcaacttggaggtagtgagaacacattca	1097
QY	1326	ccctaacacaaagaagaagatctctcccccattaaattcaactatgtttacagggaaatg	1385
Dp	1098	ccctaacacaaagaagaagatctctcccccattaaattcaactatgtttacagggaaatg	1157
QY	1386	gtaaatgtgtgctctcccggaattgtgtccaagaatgtttgaagaacccctaatcccaac	1445
Dp	1158	gtaaatgtgtgctctcccggaattgtgtccaagaatgtttgaagaacccctaatcccaac	1217
QY	1446	actgatctgaatctgtgtataaagaacacattatagccacagatgaacacccctgtggccaccccttca	1505
Dp	1218	actgatctgaatctgtgtataaagaacacattatagccacagatgaacacccctgtggccaccccttca	1277
QY	1506	gcgtgtgacgggtgagatgctgtgctgtgttcccaacaccccaagtaagacatctcagacat	1555
Dp	1278	gcgtgtgacgggtgagatgctgtgctgtgttcccaacaccccaagtaagacatctcagacat	1337
QY	1566	gaactctatctgcacaggctgtgtcaagatggagaggtatctagaaggagacatcagataagggtgc	1625
Dp	1338	gaactctatctgcacaggctgtgtcaagatggagaggtatctagaaggagacatcagataagggtgc	1397
QY	1626	tccttatgtcccccgtgctgtgaaatcccaacagcgggctatctgcgtttatbtgggctgtgggga	1685
Dp	1398	tccttatgtcccccgtgctgtgaaatcccaacagcgggctatctgcgtttatbtgggctgtgggga	1457
QY	1686	cttgaatctggaatgcttccaanaacatcaactgtgttggccaacaaagtgtgaaccaaagttaga	1745
Dp	1458	cttgaatctggaatgcttccaanaacatcaactgtgttggccaacaaagtgtgaaccaaagttaga	1517
QY	1746	tgataatgctcttcagtgcttgaagaataactatcgttatgaaggccaactatgggactg	1805
Dp	1518	tgataatgctcttcagtgcttgaagaataactatcgttatgaaggccaactatbtgggactga	1577
QY	1806	acttttgagacacactatgaagaagcgtgtgacacgtgtgtgggccaagagcatgtatacaaacatgc	1865
Dp	1578	acttttgagacacactatgaagaagcgtgtgacacgtgtgtgggccaagagcatgtatacaaacatgc	1637
QY	1866	tcaagaactgtctggagaagatgtgtgtgtgggagaacagggcttcttgacattcgtgtgacatcct	1925

D	b	1638	tcaagaactgcgagacagtgtgggtggagaccaggcttcgaatcgtggcatcc	1697
O	y	1926	taggaagaaggcgctcctttagatttgttgtaagatagatcttttccttgcacaattgct	1985
D	b	1698	taggataaagaaggctcgtctatttagattgttggtaagtagatccttttccttgcaattgct	1757
O	y	1986	gccctgggtgaalgcgtcgtctgttcctcaccccccaacctagtagtctcctcactaac	2045
D	b	1758	gctctgggtgaatgcgtcgtcttctctcaccccccaacctagtagtctcctcactaac	1817
O	y	2046	cctcactaaagtgaagaagaacactgcctgtatattggaggagatggaaggagatatgtgtg	2105
D	b	1818	cctcactaaagtgaagaagaacactgcctgtatattggaggagatggaaggagatatgtgtg	1877
O	y	2106	agcaactgtattccaagttagatgcctcctgtttaagctttccatctcttggagctgccttc	2165
D	b	1878	agcaactgtattccaagttagatgcctcctgtttaagctttccatctcttggagctgccttc	1937
O	y	2166	cctaataatccaggtttgtgtagcgttggagagaacttgatgtgaaagagaaccttcctt	2225
D	b	1938	cctaataatccaggtttgtgtagcgttggagagaacttgatgtgaaagagaaccttcctt	1997
O	y	2226	cgttactgtttaacttaaaaaataaatgcttcctcgattccaagta	2268
D	b	1998	cgttactgtttaacttaaaaaataaatgcttcctcgattccaagta	2040
 RESULT 5 AAC62134 ID AAC62134 standard; DNA; 2128 BP. XX AC AAC62134; XX DT 06-MAR-2001 (first entry) DE Nucleotide sequence of beta-1-6-N-acetylglucosaminyltransferase. XX KW Human; beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M; inflammation membrane protein; branched sialyl Lex; L-selectin; immune reaction; inflammation; tissue rejection; tumour metastasis; ss. XX OS Homo sapiens.				
FH	Key	Location/Qualifiers		
FT	CDS	354..1670		
FT		/tag= a		
FT		/product= "beta-1-6-N-acetylglucosaminyltransferase"		
FT	polyA_signal	2100..2105		
FT		/tag= b		
PN	US6136580-A.			
PD	24-OCT-2000.			
PF	19-JAN-1999;	990S-0233506.		
PR	19-JAN-1999;	990S-0233506.		
PA	(BURN-) BURNHAM INST.			
PI	Fukuda M, Yeh J;			
DR	WPI: 2001-040238/05.			
PS	P-PSDB: AAB30518.			
PT	New C2GNT-M polypeptides having core 2, core 4 and I branching			
PT	beta-1-6-N-acetylglucosaminyltransferase activities for preparing			
PT	reagents useful for diagnosing, preventing or treating inflammation or			
PT	tumour metastasis			
PS	Example 1; Fig 4; 25pp; English.			

CC The present sequence encodes a human
CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
CC and I branching activities. It is designated C2GnT-M. C2GnT-M is a
CC membrane protein that is predominantly expressed in colon, small
CC intestine, trachea, stomach and thyroid, as well as in certain cancer
CC cell lines. C2GnT-M polypeptides may be used to prepare molecules having
CC highly branched sialyl Lex and L-selectins, which may be subsequently
CC used to modulate immune reactions, e.g. inflammation and tissue
CC rejection, and to prevent or inhibit tumour metastasis.
XX
SQ Sequence 2128 BP; 569 A; 477 C; 526 G; 556 T; 0 other;

Query Match 85.8%; Score 1990.8; DB 22; Length 2128;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2004; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 244 agatattaagagagagcctgaactgtctcttgacatcttatgaagtcagaataacc 303
DB 102 agatattaagagagagcctgaactgtctcttgacatcttatgaagtcagaataacc 161
QY 304 ttcttgagaggttaagaatcagggagacatgtgttcacatttgtccacgagacacg 363
DB 162 ttcttgagaggttaagaatcagggagacatgtgttcacatttgtccacgagacacg 221
QY 364 ccagcttctcacttggaaacagaaatcacgcctgtgtgaagagatcatccctaaagcagagag 423
DB 222 ccagcttctcacttggaaacagaaatcacgcctgtgtgaagagatcatccctaaagcagagag 281
QY 424 aagctattaagagattgtctctctccacacttccctgtgtgtgtccacccgtctcc 483
DB 282 aagctattaagagattgtctctctccacacttccctgtgtgtgtccacccgtctcc 341
QY 484 catctctgtacagatgtgttcaatgtgaagagactctgcacgcgtcatcttctgtggctctg 543
DB 342 catctctgtacagatgtgttcaatgtgaagagactctgcacgcgtcatcttctgtggctctg 401
QY 544 ggcctgcataatgtcgtgcgcacgtgtgtgtcttgaaacttcttccaggttgaaagtgtgac 603
DB 402 ggcctgcataatgtcgtgcgcacgtgtgtgtcttgaaacttcttccaggttgaaagtgtgac 461
QY 604 tctacacacacttgggtctctggagttccagaggaatctcaaaagccagctactgttagaataatctg 663
DB 462 tctacacacacttgggtctctggagttccagaggaatctcaaaagccagctactgttagaataatctg 521
QY 664 tataattctctgaacttccagcaaaagaggtctatacactgttcaagggttcaaccgagag 723
DB 522 tataattctctgaacttccagcaaaagaggtctatacactgttcaagggttcaaccgagag 581
QY 724 gaaccaagaagcagtgcttcaaggtctatcttgaaataacctgtgaagtcacaagaagcagagag 783
DB 582 gaaccaagaagcagtgcttcaaggtctatcttgaaataacctgtgaagtcacaagaagcagagag 641
QY 784 ccttccacagacccacactactctctccctccacagagactgtgaagcctttaaagcgtgaa 843
DB 642 ccttccacagacccacactactctctccctccacagagactgtgaagcctttaaagcgtgaa 701
QY 844 aggaagttcatatacagttccacactgacgaagaagaggttgaggtccctattgtcactact 903
DB 702 aggaagttcatatacagttccacactgacgaagaagaggttgaggtccctattgtcactact 761
QY 904 atgtgtatctcaatgaagaagattgaaacttgaagaagcctactgcagctgtgtatgcccct 963
DB 762 atgtgtatctcaatgaagaagattgaaacttgaagaagcctactgcagctgtgtatgcccct 821
QY 964 cagaacataatactgtgtccatgagtgagagaggtcccccagaacttccaaagagcggttc 1023
DB 822 cagaacataatactgtgtccatgagtgagagaggtcccccagaacttccaaagagcggttc 881
QY 1024 aaagaacattatcttctgtctcccaatgtctctcatagccagtaagcgtgttcggtgttc 1083
DB 882 aaagaacattatcttctgtctcccaatgtctctcatagccagtaagcgtgttcggtgttc 941

QY 1084 tatgtctctctgttccagaggttgacaagctgaacctcaactgcataatgtgaagcctgtctccagagc 1143
DB 942 tatgtctctctgttccagaggttgacaagctgaacctcaactgcataatgtgaagcctgtctccagagc 1001
QY 1144 tcaagtcctgtggaataatctctctgaatacatgttgaggcaggaacttctctataaagaagaat 1203
DB 1002 tcaagtcctgtggaataatctctctgaatacatgttgaggcaggaacttctctataaagaagaat 1061
QY 1204 gcaagagatgttccagagcttccaaagatgttgaaatgtgagagataagcatgagagctcagaagta 1263
DB 1062 gcaagagatgttccagagcttccaaagatgttgaaatgtgagagataagcatgagagctcagaagta 1121
QY 1264 cctctccaaagcaaaagaaccccgctgtgaataatatacactttaggttagtgagagacacatta 1323
DB 1122 cctctccaaagcaaaagaaccccgctgtgaataatatacactttaggttagtgagagacacatta 1181
QY 1324 caacctaaaccacaagaagaagatccctccctctataatatttaactagtttacaagggaat 1383
DB 1182 caacctaaaccacaagaagaagatccctccctctataatatttaactagtttacaagggaat 1241
QY 1384 gcgtacattgtgtctccagagatttgcaccaatgttttgaagaacccctaaatcccaa 1443
DB 1242 gcgtacattgtgtctccagagatttgcaccaatgttttgaagaacccctaaatcccaa 1301
QY 1444 caactgtattgaatgtgtaaaagaacattatagcccaatgaacacactcttggccacccct 1503
DB 1302 caactgtattgaatgtgtaaaagaacattatagcccaatgaacacactcttggccacccct 1361
QY 1504 cagcgtgcaaggttgatgtgcctgtgtctgttcccaaccaccccaagtagcagatctcagac 1563
DB 1362 cagcgtgcaaggttgatgtgcctgtgtctgttcccaaccaccccaagtagcagatctcagac 1421
QY 1564 atgacttctatttgcagagctgtgtcaagttgagcaaggttcatgagagagacatcogaataaggt 1623
DB 1422 atgacttctatttgcagagctgtgtcaagttgagcaaggttcatgagagagacatcogaataaggt 1481
QY 1624 gctccttaatgtctccctgtctcttggaatcccaagcggttatctgtcttattgaggtcgtgg 1683
DB 1482 gctccttaatgtctccctgtctcttggaatcccaagcggttatctgtcttattgaggtcgtgg 1541
QY 1684 gacttgatattgagatgtcttcaaaaacacacactgtgttgccaaagaatttgaccocaaagta 1743
DB 1542 gacttgatattgagatgtcttcaaaaacacacactgtgttgccaaagaatttgaccocaaagta 1601
QY 1744 gatgataatgtctctctcagtgcttgaagaatactactagttataaaggccatctatgagact 1803
DB 1602 gatgataatgtctctctcagtgcttgaagaatactactagttataaaggccatctatgagact 1661
QY 1804 gaactttgagacaacacatataagagcgttgtctactcgtgtgggccaagacatgataaacaat 1863
DB 1662 gaactttgagacaacacatataagagcgttgtctactcgtgtgggccaagacatgataaacaat 1721
QY 1864 gctcagaactgtctggagacaggtgtgggttgagagacccgggttcgtcaactcgtgtgaccc 1923
DB 1722 gctcagaactgtctggagacaggtgtgggttgagagacccgggttcgtcaactcgtgtgaccc 1781
QY 1924 tttaagaataagaggtctgtataattagatgtgtgtgaaatagatcttctgtcccttgaaatg 1983
DB 1782 tttaagaataagaggtctgtataattagatgtgtgtgaaatagatcttctgtcccttgaaatg 1841
QY 1984 ctgcctgtgtgtaatgtcgtctgtctctcaaccctcaaccctagttagttctctccactaact 2043
DB 1842 ctgcctgtgtgtaatgtcgtctgtctctcaaccctcaaccctagttagttctctccactaact 1901
QY 2044 ttctcactaaggttgaaatgagaaactgtgtgtatagggagaggtgaagagagataatgtgt 2103
DB 1902 ttctcactaaggttgaaatgagaaactgtgtgtatagggagaggtgaagagagataatgtgt 1961
QY 2104 agagacattgatttcaattgaatgtcagctcgtgtgtagcttccactcctcgtgtgagcgtgcgt 2163
DB 1962 agagacattgatttcaattgaatgtcagctcgtgtgtagcttccactcctcgtgtgagcgtgcgt 2021
QY 2164 tcctaataattccaggttctgtgtagcgttgagagagaaacttgatgtgaaagaacacttccc 2223

Db 2022 tcctaataatccaggttggaacgtgagagagacttgtaggaagagac--ttcc 2079
Oy 2224 ttctgtactgttaacttaataata 2251
Db 2080 ttctgtactgttaacttaataata 2107

RESULT 6

AAA96569 standard; DNA; 1317 BP.

AAA96569;

08-FEB-2001 (first entry)

DNA encoding a core 2 beta-1,6-N-acetylglucosaminyltransferase.

Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLCNAc-T;
cancer; cardiovascular disorder; inflammatory disorder; asthma;
rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
septic shock; adult respiratory distress syndrome; ARDS; cancer;
platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
diverticulitis; ulcerative colitis; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 1..1317

/*tag= a /product= "core 2 beta-1,6-N-acetylglucosaminyltransferase"

CA229636-A1.

03-AUG-2000.

03-FEB-2000; 2000CA-229636.

03-FEB-1999; 99US-0118674.

(GLYC-) GLYCODESIGN INC.

Korczak B, Lew A;

WPI: 2000-594746/57.

P-PSDB; AAB18995.

New nucleic acid molecules of core 2
beta-1,6-N-acetylglucosaminyltransferase useful for providing new
compositions for treatment of disorders mediated by the enzyme -
including cancer, cardiovascular and inflammatory disorders.
Claim 4; Page 50; 66pp; English.

The present sequence encodes a human core 2

beta-1,6-N-acetylglucosaminyltransferase (core2b GLCNAc-T) polypeptide.

The polypeptide can be used to treat diseases and disorders, such as
cancer, cardiovascular disorders and inflammatory disorders including
asthma, rheumatoid arthritis, inflammatory bowel disease,

arteriosclerosis, septic shock, adult respiratory distress syndrome

(ARDS) and cancer. Various platelet-mediated pathologies such as

atherosclerosis and clotting can also be treated. The polypeptides of

the invention are predominantly expressed in gastrointestinal tissue

(stomach, colon, intestine, testis) and are elevated in cancer.

Gastrointestinal disorders that may be prevented or treated include

ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis

and ulcerative colitis. The antibodies may be used in

immuno-histochemical analysis, to detect the novel polypeptide and to

localize it to particular cells and tissues and to specific subcellular

locations and to quantitate the level of expression.

Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;

Query Match 56.8%; Score 1317; DB 21; Length 1317;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 496 atggttcaatggaagagctctgcagctgcgaactcttggtggtcctggtcgtatgt 555
Db 1 atggttcaatggaagagctctgcagctgcgaactcttggtggtcctggtcgtatgt 60
Oy 556 ctgctgcagcactgtgctctgaactctcttcaggttgaagtgtgactgtaccactgt 615
Db 61 ctgctgcagcactgtgctctgaactctcttcaggttgaagtgtgactgtaccactgt 120
Oy 616 ggtctgaggtccaggaatctcaagccaagtctgtgaagaatcttgtataattctgt 675
Db 121 ggtctgaggtccaggaatctcaagccaagtctgtgaagaatcttgtataattctgt 180
Oy 676 aaacttccagcaagaagttctacactgttcaagggttcaaccggagggaaccaagagca 735
Db 181 aaacttccagcaagaagttctacactgttcaagggttcaaccggagggaaccaagagca 240
Oy 736 gtgtctcaggtatcttgaataacctggaggtcaagaaagcgaagaccttccacagac 795
Db 241 gtgtctcaggtatcttgaataacctggaggtcaagaaagcgaagaccttccacagac 300
Oy 796 acccaactacctccctcaaccagagactgtgagcacttcaaggtcgaagaagatcata 855
Db 301 acccaactacctccctcaaccagagactgtgagcaacttcaaggtcgaagaagatcata 360
Oy 856 cagttcccacttgagcaagaaggtgtgagttccctatgtgacatctatgttgtatcat 915
Db 361 cagttcccacttgagcaagaaggtgtgagttccctatgtgacatctatgttgtatcat 420
Oy 916 gagaagattgaaaactttgaaaggctactgcagctgtgtatgtccctcagaacataac 975
Db 421 gagaagattgaaaactttgaaaggctactgcagctgtgtatgtccctcagaacataac 480
Oy 976 tgtgtccatgttgatgagaaggtccccaagaacttccaagaaggcgttcaagaactatt 1035
Db 481 tgtgtccatgttgatgagaaggtccccaagaacttccaagaaggcgttcaagaactatt 540
Oy 1036 tcttgctcccaaatgctctcaacagtaagcagtggttcggtgttatctccctcgg 1095
Db 541 tcttgctcccaaatgctctcaacagtaagcagtggttcggtgttatctccctcgg 600
Oy 1096 tccaaggtgcaagctgacctcaactgcatggaagactgtgtccagagctcagtcgctg 1155
Db 601 tccaaggtgcaagctgacctcaactgcatggaagactgtgtccagagctcagtcgctg 660
Oy 1156 aaatcttccgaatacatgtggaagcagacttctctataagaagaaatgcagagtgctc 1215
Db 661 aaatcttccgaatacatgtggaagcagacttctctataagaagaaatgcagagtgctc 720
Oy 1216 caggtctcaaatgttgaatggaagataagcatgagctcagagtaacctctcaagac 1275
Db 721 caggtctcaaatgttgaatggaagataagcatgagctcagagtaacctctctcaagac 780
Oy 1276 aaagaaccgcgtggaatatcatcttgaagtagtgagagacacattaaccttaaccaac 1335
Db 781 aaagaaccgcgtggaatatcatcttgaagtagtgagagacacattaaccttaaccaac 840
Oy 1336 aagaagaaggttcccccctataatataactatgatttaagggaaatgagagactgtg 1395
Db 841 aagaagaaggttcccccctataatataactatgatttaagggaaatgagagactgtg 900
Oy 1396 gttcccgagattgttccaacatgttgaagaaccttaatacccaacactgtgttaa 1455
Db 901 gttcccgagattgttccaacatgttgaagaaccttaatacccaacactgtgttaa 960
Oy 1456 tgggtaaagaactatagaccagatgaacacctgtggtccacacctcagcgtgacgg 1515

DB 961 ttggtataaagacacttagccagatgataacacctctgggccaccttcaagctgcaagg 1020
QY 1516 tggatgtcctggtctgttcccaacaccccaagtagacatctcagaatgacttctatt 1575
DB 1021 ttggtgtcgtctgttcccaacaccccaagtagacatctcagaatgacttctatt 1080
QY 1576 gccaggtctgtatgaagtgacaggtgataagagagacatcagatggtctccttattgct 1635
DB 1081 gccaggtctgtatgaagtgacaggtgataagagagacatcagatggtctccttattgct 1140
QY 1636 cccgtgcttggaatccacccagcggtatctgtcgtttatggggtctgggacttgaattgg 1695
DB 1141 cccgtgcttggaatccacccagcggtatctgtcgtttatggggtctgggacttgaattgg 1200
QY 1696 atgcttcaaaaaccatcacctgttggccaacaagtgtgacccaagaagtagatgataatgct 1755
DB 1201 atgcttcaaaaaccatcacctgttggccaacaagtgtgacccaagaagtagatgataatgct 1260
QY 1756 cttaagtgtctagaagaatactacgttataaaggccatctatgggactgaaacttga 1812
DB 1261 cttaagtgtctagaagaatactacgttataaaggccatctatgggactgaaacttga 1317

RESULT 7

AAA96375
ID AAA96575 standard; DNA; 1221 BP.

XX AAA96575;

XX 08-FEB-2001 (first entry)

DE Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNA.

XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLCNAc-T;
KM cancer; cardiovascular disorder; inflammatory disorder; asthma;
KM rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KM septic shock; adult respiratory distress syndrome; ARDS; cancer;
KM platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KM clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KM diverticulitis; ulcerative colitis; ss.

XX Homo sapiens.

XX CA2296936-A1.

XX 03-AUG-2000.

PE 03-FEB-2000; 2000CA-2296936.

XX 03-FEB-1999; 99US-0118674.

PA (GLYC-) GLYCODESIGN INC.

PI Korczak B, Lew A;

DR WPI; 2000-594746/57.

XX New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
PS Claim 4; Page 54; 66pp; English.

XX The present sequence encodes a fragment of a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GLCNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC atherosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue

CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 1221 BP; 342 A; 285 C; 295 G; 299 T; 0 other;

Query Match 52.7%; Score 1221; DB 21; Length 1221;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 ttgaagtgtactctgtaccacttgggtctgagtcctcgggaatcctcaaacgactgct 651
DB 1 ttgaagtgtactctgtaccacttgggtctgagtcctcgggaatcctcaaacgactgct 60
QY 652 aggaatactctgtataattctcctgaacttcacgaagaaggtctalcaactgttcaagg 711
DB 61 aggaatactctgtataattctcctgaacttcacgaagaaggtctalcaactgttcaagg 120
QY 712 gtcaaccgagggagccaagaagggcagtgcttcaagctattctgaataacctggaggtcaag 771
DB 121 gtcaaccgagggagccaagaagggcagtgcttcaagctattctgaataacctggaggtcaag 180
QY 772 aagaagcgagagcctttccacagacacccactactctcctccacagagactgtgacac 831
DB 181 aagaagcgagagcctttccacagacacccactactctcctccacagagactgtgacac 240
QY 832 ttcaagctgaagaaggttcaatacagttcccaactgagcaagaagaaggtgagttccct 891
DB 241 ttcaagctgaagaaggttcaatacagttcccaactgagcaagaagaaggtgagttccct 300
QY 892 atggcatactatggtgatctcctgagaagaattgaaactttgaaggtcactggcagct 951
DB 301 atggcatactatggtgatctcctgagaagaattgaaactttgaaggtcactggcagct 360
QY 952 gtgatgccccctcagaacataactgtgtcatgctgtagtgaaggtcccaagaacttc 1011
DB 361 gtgatgccccctcagaacataactgtgtcatgctgtagtgaaggtcccaagaacttc 420
QY 1012 aaagaagcggtcacaagaatattctctgtctcccaatgcttccatagccagtaagctg 1071
DB 421 aaagaagcggtcacaagaatattctctgtctcccaatgcttccatagccagtaagctg 480
QY 1072 gtccgggtgtttatgtcctctgtccagggttgcaaggtgacactcaactcagtaagaagc 1131
DB 481 gtccgggtgtttatgtcctctgtccagggttgcaaggtgacactcaactcagtaagaagc 540
QY 1132 ttgctccagagctcaagtccgttggaataactcctgtaatacatgttggagcgacttctc 1191
DB 541 ttgctccagagctcaagtccgttggaataactcctgtaatacatgttggagcgacttctc 600
QY 1192 ataaagagcaatgcagagatgttccaggtctcaagaatgtttgaatggaggaatagcatg 1251
DB 601 ataaagagcaatgcagagatgttccaggtctcaagaatgtttgaatggaggaatagcatg 660
QY 1252 gattcagaaggtactcctctaagcacaagaacccgcctggagaatatcacttggagtgatg 1311
DB 661 gattcagaaggtactcctctaagcacaagaacccgcctggagaatatcacttggagtgatg 720
QY 1312 agagacacattacaccttaacacaagaagaagaagatcctcccttataatttaattatg 1371
DB 721 agagacacattacaccttaacacaagaagaagaagaagatcctcccttataatttaattatg 780
QY 1372 ttcaagaggaatgctacatgttgcttcccgagatttgttccacaatgtttgaagaac 1431
DB 781 ttcaagaggaatgctacatgttgcttcccgagatttgttccacaatgtttgaagaac 840
QY 1432 ccttaattcccaaacacatgattgaatggtaagaagacacttatggccagtgtaaccctc 1491

Db 841 cctaaatccacaacatgatitgaatgggtaaaagaacactatagccagatgaacaccc 900
QY 1492 tgggaccaccttcacgycgacggtgatggcctgtgctgttcccaaccacccaagtac 1551
Db 901 tgggaccaccttcacgycgacggtgatggcctgtgctgttcccaaccacccaagtac 960
QY 1552 gacatctcagacatcactctatctccaggtcgtcgaattggcaggtcattagggagac 1611
Db 961 gacatctcagacatcactctatctccaggtcgtcgaattggcaggtcattagggagac 1020
QY 1612 atcgataaagggtcctcctatgtctccctgtccttggaatccacagcggtatctcgctt 1671
Db 1021 atcgataaagggtcctcctatgtctccctgtccttggaatccacagcggtatctcgctt 1080
QY 1672 tatgggcttgggacttgatgtgatgtgcttcaaacatcactcttggccaacagttt 1731
Db 1081 tatgggcttgggacttgatgtgatgtgcttcaaacatcactcttggccaacagttt 1140
QY 1732 gaccacaaggtagatgaatactgtctcgaagtcttagaagaatacctacgtataagacc 1791
Db 1141 gaccacaaggtagatgaatactgtctcgaagtcttagaagaatacctacgtataagacc 1200
QY 1792 atctatgggactgaacttga 1812
Db 1201 atctatgggactgaacttga 1221

RESULT 8
AAVS9800
ID AAVS9800 standard; DNA; 997 BP.
XX
AC AAVS9800;
XX
DT 19-JAN-1999 (first entry)
XX
DE Human secreted protein gene 171 clone HTEBJ34.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9839448-A2.
XX
PD 11-SEP-1998.
XX
PE 06-MAR-1998; 98WO-US04493.
XX
PR 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.

PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047490.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047616.
PR 23-MAY-1997; 97US-0047617.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.

PR 22-AUG-1997; 9705-0056909.
 PR 22-AUG-1997; 9705-0056910.
 PR 22-AUG-1997; 9705-0056911.
 PR 05-SEP-1997; 9705-0057650.
 PR 05-SEP-1997; 9705-0057659.
 PR 05-SEP-1997; 9705-0057761.
 PR 12-SEP-1997; 9705-0058785.
 PR XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
 PI Reng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
 PI Kyaw H, Little DM, Li Y, Moore PA, NI J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 DR WPI: 1998-506364/43.
 DR F-PSDB: AAM75014.
 XX
 PT New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 PS
 PS Claim 1: Page 530-531. 721pp: English.
 XX
 CC This sequence represents a nucleic acid molecule designated Gene 171
 CC from the human cDNA clone HTEDJ34 (deposited as clone ATCC 97904 and
 CC ATCC 209050) which encodes a secreted human protein. The gene can be
 CC used to generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
 CC the fused protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 186
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV59511 for described uses).
 CC XX
 SO Sequence 997 BP; 275 A; 213 C; 243 G; 262 T; 4 other:

Db 361 tatctgcgtttatgaggctgaggacttggaattggatgcttcaaacatcaactgctgttgc 420
 QY 1722 caacaagtttgaccacaagtagatgataatgctcttcagctgttgtagaagaataactacg 1781
 Db 421 caacaagtttgaccacaagtagatgataatgctcttcagctgttgtagaagaataactacg 480
 QY 1782 ttataagccatcatatgagacttggaactttgagacacactatgagagcgttgcctactg 1841
 Db 481 ttataagccatcatatgagacttggaactttgagacacactatgagagcgttgcctactg 540
 QY 1842 gggcaagagatgataaacaatgctcagaacttgcctgggacagatggtggtagagccag 1901
 Db 541 gggcaagagatgataaacaatgctcagaacttgcctgggacagatggtggtagagccag 600
 QY 1902 ggccttgcaattgctgacatccttccttagaagaaggcgtcctatagatgtgtgaagt 1961
 Db 601 ggccttgcaattgctgacatccttccttagaagaaggcgtcctatagatgtgtgaagt 660
 QY 1962 agatcttgccttgcaaatgctgctgggtggaatgctgtctctacccctaac 2021
 Db 661 agatcttgccttgcaaatgctgctgggtggaatgctgtctctacccctaac 720
 QY 2022 cctagtagtccctcccaacttctcactaagtgaagaaggaactgctgtatagga 2081
 Db 721 cctagtagtccctcccaacttctcactaagtgaagaaggaactgctgtatagga 780
 QY 2082 gaatgaagagagatagatgtgtgtagaacttgatcagttgactgctgtgtgtagct 2141
 Db 781 gaatgaagagagatagatgtgtgtagaacttgatcagttgactgctgtgtgtagct 840
 QY 2142 ttccattcgtgtagctgctgcttccataatccaggttggtagcgtgtagaagaact 2201
 Db 841 ttccattcgtgtagcgtgcttccataatccaggttggtagcgtgtagaagaact 900
 QY 2202 ttgatgaaagagaacttccctctcctgtagtgttaactttaaataatagccctgatt 2261
 Db 901 ttgatgaaagagaacttccctctcctgtagtgttaactttaaataatagccctgatt 960
 QY 2262 caaagta 2268
 Db 961 caaagta 967

RESULT 9
 AAV59681
 ID AAV59681 standard; DNA: 777 BP.
 XX
 AC AAV59681:
 DT 19-JAN-1999 (first entry)
 XX
 DE Human secreted protein gene 171 clone HTEDJ34.
 XX
 KW Human: secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclasts; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 OS Homo sapiens.
 XX
 PN MO9839448-A2.
 XX
 PD 11-SEP-1998.
 XX
 PF 06-MAR-1998; 98WO-US04493.
 XX
 PR 02-OCT-1997; 97US-0061060.
 PR 07-MAR-1997; 97US-0038621.
 PR 07-MAR-1997; 97US-0040161.

PR	07-MAR-1997	97US-00401162
PR	07-MAR-1997	97US-00401163
PR	07-MAR-1997	97US-00403033
PR	07-MAR-1997	97US-00403034
PR	07-MAR-1997	97US-00403036
PR	07-MAR-1997	97US-00406826
PR	11-APR-1997	97US-00433311
PR	11-APR-1997	97US-00433312
PR	11-APR-1997	97US-00433313
PR	11-APR-1997	97US-00433314
PR	11-APR-1997	97US-00433568
PR	11-APR-1997	97US-00433569
PR	11-APR-1997	97US-00433576
PR	11-APR-1997	97US-00433578
PR	11-APR-1997	97US-00433580
PR	11-APR-1997	97US-00433674
PR	11-APR-1997	97US-00436792
PR	23-MAY-1997	97US-00475801
PR	23-MAY-1997	97US-00475802
PR	23-MAY-1997	97US-00475803
PR	23-MAY-1997	97US-00475812
PR	23-MAY-1997	97US-00475821
PR	23-MAY-1997	97US-00475828
PR	23-MAY-1997	97US-00475834
PR	23-MAY-1997	97US-00475845
PR	23-MAY-1997	97US-00475854
PR	23-MAY-1997	97US-00475861
PR	23-MAY-1997	97US-00475869
PR	23-MAY-1997	97US-00475871
PR	23-MAY-1997	97US-00475873
PR	23-MAY-1997	97US-00476114
PR	23-MAY-1997	97US-00476115
PR	23-MAY-1997	97US-00476116
PR	23-MAY-1997	97US-00476117
PR	23-MAY-1997	97US-00476118
PR	23-MAY-1997	97US-00476332
PR	23-MAY-1997	97US-00476333
PR	06-JUN-1997	97US-00483634
PR	06-JUN-1997	97US-00483635
PR	13-JUN-1997	97US-00483714
PR	08-JUL-1997	97US-00496100
PR	16-JUL-1997	97US-00519266
PR	18-AUG-1997	97US-00552724
PR	22-AUG-1997	97US-00556330
PR	22-AUG-1997	97US-00556331
PR	22-AUG-1997	97US-00556332
PR	22-AUG-1997	97US-00556336
PR	22-AUG-1997	97US-00556337
PR	22-AUG-1997	97US-00556662
PR	22-AUG-1997	97US-00556664
PR	22-AUG-1997	97US-00556645
PR	22-AUG-1997	97US-00556652
PR	22-AUG-1997	97US-00556654
PR	22-AUG-1997	97US-00556674
PR	22-AUG-1997	97US-00556675
PR	22-AUG-1997	97US-00556676

PR	22-AUG-1997	97US-0056877
PR	22-AUG-1997	97US-0056878
PR	22-AUG-1997	97US-0056879
PR	22-AUG-1997	97US-0056880
PR	22-AUG-1997	97US-0056881
PR	22-AUG-1997	97US-0056882
PR	22-AUG-1997	97US-0056884
PR	22-AUG-1997	97US-0056886
PR	22-AUG-1997	97US-0056887
PR	22-AUG-1997	97US-0056888
PR	22-AUG-1997	97US-0056889
PR	22-AUG-1997	97US-0056892
PR	22-AUG-1997	97US-0056893
PR	22-AUG-1997	97US-0056894
PR	22-AUG-1997	97US-0056903
PR	22-AUG-1997	97US-0056908
PR	22-AUG-1997	97US-0056909
PR	22-AUG-1997	97US-0056910
PR	22-AUG-1997	97US-0056911
PR	05-SEP-1997	97US-0057650
PR	05-SEP-1997	97US-0057669
PR	05-SEP-1997	97US-0057761
PR	12-SEP-1997	97US-0058785
PR	12-SEP-1997	97US-0058786

(HUMA-) HUMAN GENOME SCI INC.

Bednarik D, Brewer LA, Carter KC, Dunn R, Ebner R, Endress GA,
Feng P, Ferlie AM, Florence CL, Greene JM, Hu JS,
Kyaw H, Laflair DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z,
WPI: 1998-506364/43.
P-PSDB: AAN74898.

New isolated human genes and the secreted polypeptide(s) they encode - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders

Claim 1; Page 410-411; 721pp; English.

This sequence represents a nucleic acid molecule designated Gene 171 from the human cDNA clone HTD3D34 (deposited as clone ATCC 97904 and ATCC 209050) which encodes a secreted human protein. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of the fused protein as compared to the human protein only.

The invention relates to 186 novel genes and their fragments (nucleic acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W5026) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 186 polynucleotides, based on which tissues they are most highly expressed in (see AAV59511 for described uses).

Sequence 777 BP; 203 A; 142 C; 194 G; 198 T; 40 other;

Query Match	21.0%;	Score 486.2;	DB 19;	Length 777;
Best Local Similarity	81.9%;	Pred. No. 2.3e-145;		
Matches 610;	Conservative 34;	Mismatches 67;	Indels 34;	Gaps 7

QY	1558	tcaacatagactctatctgacagcgtcgtgccaagtgcaggtgcatagtatgagagacatcgat	1612
	8	kcaagacatgactctatctgacagcgtcgtgccaagtgcaggtgcatagtatgagagacatcgat	67
QY	1618	aaggtctccttatctcctctgctctggaatccaccacagcggtcatctgcgcttatgg	1677
	68	aaggtggtcctctatgctcctcgtccttcggaatccaccacagcggtcatctgcgcttatgg	127
QY	1678	gtctggagacttgatctgagcttccaagaacatcacctcttgcgcaacaagtcttaccca	1737

D 128 gctgggactagaaattggaatgtcctcaaaaacacatcactglttggcacaagttagacca 187
Q 1738 aaggtatagataatgtctctcaatgtcttagaagaatactactgattataagcc-----at 1793
D 188 aaggtatagataatgtctctcaatgtcttagaagaatactactgattataagggccattct 247
Q 1794 ctatggactgaactttagacacactat-----gagagcgttgcatacc 1837
D 248 attggacactgaactttagaagacacamatltgaagagcgcttgcataccgcttggg99gc 307
Q 1838 tgtgggcaagagacatgtacaaacatgtcgaactgtgtggagc--agtgtgggtggga 1895
D 308 caaagagacatgtacaaacatgttgcgycargaamyltgggkkgamcarkkkkgkkggga 367
Q 1896 gaccagggccttgcaa--ttcgtggacatcctta--gataaagagggctgattagat 1950
D 368 rrmcmrgggttgcacaaatcckkgcmwccytttaggttaarrggckltwatlgat 427
Q 1951 tgtgggtaaa--agtagatcttggccttgcaaatgtcgtcgtgggtgaatgtcgtgt 2006
D 428 tgtgggtaaaagtagatcttggccttgcaaatgtcgtcgtgggtgaatgtcgtgt 487
Q 2007 tctctc--acccttaaccctagtagttcctcctcaacttctcactcaatggagaatga 2063
D 488 tctctcctcaccctcaacccatagtagttcctcctcaacttctcactcaatggagaatga 547
Q 2064 gaactgtgtatagagagagatgtgaagagagatgtgttagagacactgttccagltg 2123
D 548 gaactgtgtatagagagagatgtgaagagagatgtgttagagacactgttccagltg 607
Q 2124 aatgcctgtgtatagagagagatgtgttagagagagatgtgttagagagagatgtgt 2183
D 608 aatgcctgtgtatagagagagatgtgttagagagagatgtgttagagagagatgtgt 667
Q 2184 gtagcgt 2243
D 668 gtagcgt 727
Q 2244 aataatagctcctgattcaagaatga 2268
D 728 aataatagctcctgattcaagaatga 752
RESULT 10
AAA96571
ID AAA96571 standard; DNA; 524 BP.
AC AAA96571:
XX
DT 08-FEB-2001 (first entry)
XX
DE Noncoding region of core 2 beta-1,6-N-acetylglucosaminyltransferase gene.
XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLCNAC-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.
XX
OS Homo sapiens.
XX
PN CA2296936-A1.
XX
PD 03-AUG-2000.
XX
PE 03-FEB-2000; 2000CA-2296936.
XX
PR 03-FEB-1999; 99US-0118674.
XX
PA (GLYC-) GLYCODESIGN INC.
XX

PI Kozczak B, Lew A;
XX
XX WPI; 2000-594746/57.
DR
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
PS Disclosure; Page 52; 66pp; English.
XX
CC The present sequence represents a noncoding region from a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GLCNAC-T) gene.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 524 BP; 171 A; 90 C; 127 G; 136 T; 0 other;

Query Match 19.7%; Score 456; DB 21; Length 524;
Best Local Similarity 100.0%; Pred. No. 9,5e-136;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q 1813 gacacataatgagagcgttctactctgttgggcaagagcatgatacaaatgctcagaac 1872
D 1 gacacacatagagagcgttctactctgttgggcaagagcatgatacaaatgctcagaac 60
Q 1873 ttgtctggagacagtgttgggtggagacagagcgttgaatctgtgagatcctttagata 1932
D 61 ttgtctggagacagtgttgggtggagacagagcgttgaatctgtgagatcctttagata 120
Q 1933 agagagcgtctatagattgtgggtgaagtagatcttctgcttgcgaattgctcgtgg 1992
D 121 agagagcgtctatagattgtgggtgaagtagatcttctgcttgcgaattgctcgtgg 180
Q 1993 tgaatgcgtctgttctcctcaccctcaacctagtagtctcctcaacttctcacta 2052
D 181 tgaatgcgtctgttctcctcaccctcaacctagtagtctcctcaacttctcacta 240
Q 2053 agtgaagatgagacatgtctgtatagagagatgtgaagagagagatgtgttagaacact 2112
D 241 agtgaagatgagacatgtctgtatagagagatgtgaagagagagatgtgttagaacact 300
Q 2113 gattcagttgaatgctcgtcgtgtagtcttccattctgtgagatcgcgttctataa 2172
D 301 gattcagttgaatgctcgtcgtgtagtcttccattctgtgagatcgcgttctataa 360
Q 2173 ttcaagtttgtagcgttgggagagaaacttgaaggagagaaacttcccttctgact 2232
D 361 ttcaagtttgtagcgttgggagagaaacttgaaggagagaaacttcccttctgact 420
Q 2233 gtaacttaaaataatagctcctgattcaagaatga 2268
D 421 gtaacttaaaataatagctcctgattcaagaatga 456
RESULT 11
AAX24043
ID AAX24043 standard; cDNA; 2204 BP.
AC AAX24043:
XX
XX

KW Human; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
 KW UDP-GlcNAc:Galbeta1-3GalNAc:alpha1-6
 KW N-acetylglucosaminyltransferase; ss.
 XX
 OS Homo sapiens.
 XX
 XX US6131578-A.
 PN
 PD 17-OCT-2000.
 PD
 PF 02-OCT-1997; 97US-0943058.
 PF
 PR 02-OCT-1996; 96US-0046876.
 PR
 XX (KING/) KING G L.
 PA (NISH/) NISHIO Y.
 PA (KOYA/) KOYA D.
 PA (DENN/) DENNIS J W.
 PA (WARR/) WARREN C E.
 XX
 PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
 PI
 DR MPI; 2000-678642/66.
 DR P-PSDB; AAB30298.
 XX
 XX Preventing or treating cardiomyopathy associated with diabetes mellitus
 PT and hyperglycemia, comprises administering a substance that inhibits
 PT core 2 N-acetylglucosamine-T activity -
 XX
 XX
 PS Claim 3; Fig 9; 21pp; English.
 XX
 CC The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc
 CC alpha1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
 CC and coding sequences. The enzyme is associated with cardiomyopathy in
 CC diabetes and hyperglycaemia sufferers. The invention also provides
 CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
 CC in treatment.
 CC
 SO Sequence 2204 BP; 641 A; 414 C; 498 G; 651 T; 0 other;

Query Match 15.9%; Score 369.4; DB 21; Length 2204;
 Best Local Similarity 60.2%; Pred. No. 1.7e-107;
 Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 772 aagaagcgaagccttcacagacacccactacccctccacagagactgtgagcac 831
 II IIII I IIII IIII I IIII IIII
 DB 490 aaaaagcgccctcggtgacacccgagactatataaacatgaccagtgctctct 549
 QY 832 ttcaagcgtaaaggaagttcacaagttccactgagcaagaagagtggtggtccct 891
 IIII I IIII I IIII I IIII IIII IIII
 DB 550 ttcatcaagagacgcaaatatattgtagaaccccttagtaagaagagcgagattcca 609
 QY 892 attgacactatgtggtatcatgagaagattgaaaactttgaaagctactgcagct 951
 II IIII IIII IIII IIII IIII IIII IIII IIII
 DB 610 atagcatattcttaagtggtcaccacaaagatgaaatgcttgacagcgctgagggcc 669
 QY 952 gtgtatgccctcagaacatactgttcacatgtgagatgagaagttccccaagacttc 1011
 IIII IIII IIII I IIII IIII IIII IIII
 DB 670 atcataatgctcagaattcttctgtcatatgttgacaaatcctcgagatctctat 729
 QY 1012 aaagagcggtcacaagaattattctgtctcccaaatgtcttcacagcaagtaagctg 1071
 II IIII I IIII I IIII I IIII IIII IIII
 DB 730 ttgactgcaagtgtgacatgccttcctgttttagtaagtctttgtgacagcgattg 789
 QY 1072 gtctcggtgtttatgctctctgttcagagtgcaagctgaactccaactctatgagaac 1131
 I IIII IIII IIII I IIII I IIII IIII IIII
 DB 790 gagagtggtgttatgcatcgtgagcggtgtcagctgacccaactctcatgaaagat 849
 QY 1132 ttgctcagagcgcacagtcgacgtaaatctctgatacatgttgagcgagacttccct 1191
 I IIII I IIII I IIII I IIII IIII IIII
 DB 850 ctctatgcaatgagtgcaaacactggaagtaactgtgataaactcttggtatgatttccc 909

QY 1192 ataaagacaaatgcaagatggtccaggtctcgaagtgttgaaatggtggaatagcatg 1251
 II IIII IIII IIII IIII IIII IIII IIII
 DB 910 attaaacacacctaagaattgttcaggaagctcaagtgttaattggaagaacaacactg 969
 QY 1252 gactcagagttactcctctaagcacaagaagaaccgcctcggaatataactttggagtagt 1311
 II IIII IIII IIII IIII IIII IIII IIII
 DB 970 gaaacggaagagatgcatcccataaagaagaaggggaagaagcgtatgaggtcgtt 1029
 QY 1312 agagacacattacacccaacaagaagaagatccctccctataattataatg 1371
 I IIII I IIII I IIII I IIII I IIII
 DB 1030 aatggaagagctgcaacaacacagggacgtcaaaaatgtctctccatccagcaaacctctc 1089
 QY 1372 tttaacaggaatgctgacattgtgcttcccgagatttgcctcaacatgttttgaagaac 1431
 II IIII IIII IIII IIII IIII IIII IIII
 DB 1090 ttcttgcagtgctcactctcgtgtcagtaggaagtatgtggatgtatgtactacaagaat 1149
 QY 1432 ccttaatcccaacaacactgattggttaagtagacacttatagcccaatgtagaacctc 1491
 IIII IIII I IIII I IIII I IIII I IIII
 DB 1150 gaaaaaatcccaaaaagtgtgagagtggtgcacaagacacacacgacctgtatgatatc 1209
 QY 1492 tgggccaaccttcagcgtgacagtggtatgctgtctgttcccaaccacccaagttac 1551
 IIII IIII I IIII I IIII I IIII I IIII
 DB 1210 tgggccaaccttcacaagaatctcctgaagctccggcctcactccctgcagcacaagat 1269
 QY 1552 gacatctcaaacatgactttatgccaagctgtgtcaagtgtgcaaggtctatggaagac 1611
 II IIII IIII IIII IIII IIII IIII IIII
 DB 1270 gatctatctgacatgcaagaagcagtggttccaggtgttcaagtgtgcaagtttgaagtgat 1329
 QY 1612 atcgaataggtgtctcttatgtctcctgtcctggaatccacacagcggtctatcgtgtt 1671
 I IIII IIII IIII I IIII IIII IIII IIII
 DB 1330 gtttccaaggtgtctcctcctcccgccctgcgagtaggaatccatgtgcgtcgaagtgtcat 1389
 QY 1672 tatggagctggtgagacttgaaattgtagtcttcaaaacacatcactgttggccaagaatt 1731
 II IIII IIII IIII IIII IIII IIII IIII
 DB 1390 ttcgagctgtgtgactgtaactgtagtctgtgcgaacacacactgtttgccaataagtt 1449
 QY 1732 gaccacaaggtatgatatgtcttcttcagtgctttagaagaataacactgaattataagc 1790
 IIII IIII I IIII I IIII I IIII I IIII
 DB 1450 gacgtgtatgttgaaccttctgcatccatccagttgtttagatgagatttgagacacaagc 1508

RESULT 13
 AA061559 standard; cDNA; 2105 BP.
 ID AA061559
 XX
 AC AA061559;
 XX
 DT 07-OCT-1994 (first entry)
 XX
 DE cDNA sequence of human core 2 beta 1-6
 DE N-acetylglucosaminyltransferase (C2Gnt or core 1-6 AGT).
 XX
 KW C2Gnt; 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;
 KW O-glycan; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 220..1504
 FT /*tag= a
 FT polyA_signal 1913..1918
 FT /*tag= b
 FT misc_signal 248..314
 FT /*tag= c
 FT /*label= signal/membrane-anchoring domain
 PN BP590747-A.
 PD 06-APR-1994.
 PD
 PD 29-SEP-1993; 93EP-0250268.
 PF
 XX
 XX 01-OCT-1992; 92US-0955041.
 PR

XX This invention describes a method for screening for a substance that
 CC may be used to prevent or treat cardiomyopathy associated with diabetes
 CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
 CC (UDP-GlcNAc:Galbeta1-3GalNAc-alpha beta-1,6-N-acetylglucosaminyl-
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in
 CC the presence of a test substance under conditions whereby the core 2
 CC GlcNAc-T produces a reaction product, determining the amount of reaction
 CC product, and comparing the amount of reaction product with the amount of
 CC reaction product in the absence of the test substance, where lower amounts of
 CC reaction product in the presence of the test substance indicate that the
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
 CC methods for preventing or treating cardiomyopathy associated with
 CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
 CC germ cells and somatic cells all contain a DNA construct introduced into
 CC the animal or an ancestor of the animal at an embryonic stage, where
 CC incorporation of the DNA construct into the germ line of the animal
 CC causes the animal to develop cardiomyopathy similar to that associated
 CC with diabetes mellitus and hyperglycaemia. This sequence encodes the rat
 CC DH1 protein which is used to describe the method of the invention.
 XX

SQ Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match 15.7%; Score 364.6; DB 19; Length 5010;
 Best Local Similarity 58.3%; Pred. No. 1e-105;
 Matches 659; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

QY 686 caaagaagctatcaactgttcagggtccacccaggagcccaagagcagtgcttagg 745
 Db 959 ccataagcaatgttaattgaccacaaagtttacagggtgaccacgaagaatccagaag 1018
 QY 746 ctattcgaataacacctggaaggtcaa--gagaagcgaagagcccttaccagacaccac 802
 Db 1019 tgaagcttgagatcaactaacgltgcaatcaagaagcgctccgagcgacacccatgact 1078
 QY 803 acctccctccacagaagactgtgagcacttcaagctgaagaagatcatatcagttcc 862
 Db 1079 atataaacatgaccggaactgcgcctctctcatcaggaacacgaatatataatgagc 1138
 QY 863 cactgagcaagaagaaggtgaggttccctatctgcatactatggtatcatatgagaga 922
 Db 1139 cccttactaagaagaaggttggtcttccaatgtcatalccatagtgttcatatataaga 1198
 QY 923 ttgaaaactttgaaagctactctgagcgtgtgtatgcccctccacaataataatgttcc 982
 Db 1199 ttgacatgcttgacaagctcctcctgagggccatctatagccacagaatttctacgattc 1258
 QY 983 atgtgagtgagaaggtcccccagaacttcaagaagcggtcacaagaatattcttctgt 1042
 Db 1259 acgttgacagaagaagcagaagatccttlttagccggtgcagggcatgtgacatcgtct 1318
 QY 1043 tcccaaatgtctcatagccagtaagctgtgttcgggtgttatgctcctcgtgccagg 1102
 Db 1319 ttgataatgtcttctgtgcccagccaglttgagagagtgltgtatataagctccggagtc 1378
 QY 1103 tgcagcgtgacctcaactgtgaagaactgtgtccagactcgaagtcagctgtggaatact 1162
 Db 1379 ttaaggtgcctcaactgatgaagaactgttccagaaatgaatgacaaactgtgagact 1438
 QY 1163 tcttgataatagtgtgagcagacttctctataaagacaaatgcagagatgtccagctc 1222
 Db 1439 tgatcaatcttctgtgtatgtattcccttataaaccaaccctgtgaattgttcagagc 1498
 QY 1223 tcaagatgttgatggaggaatagcaltgagagtcagaggtacccctcaagaacaaagaa 1282
 Db 1499 tcaagtccttcacagggaaacagcctggaagaactggaagatgctcccaacaagaag 1558
 QY 1283 cccgcttgaaatactactgtgagtgatgtgagacacattacacccaacaagaaga 1342
 Db 1559 agaggttgaaaaaaacgltacacggtgtgtgagcggaaagcgcagacaacacggaggtgtca 1618
 QY 1343 aggatcctcccttataatataatglttatacagggaatgcgtacattgtgcttcc 1402

Db 1619 aggcgcagctccactcaactcaactcctctcttcttcaagcagcgctatttctgttca 1678
 QY 1403 gaeatttgcctcaacatgatttgaagacccttaaccctcaacactgattgaatggttaa 1462
 Db 1679 gggaaatgtgaagctatgtctgtgaaataaataatattcaaaagttcatgaaatgggac 1738
 QY 1463 aagacattatagcccaatgaaacacccctcgtggccaccccttcagcgtgtcaggtgagtc 1522
 Db 1739 aggaacatacagcccgatgagttcctctggccaccacccaagaagatccctgaagtc 1798
 QY 1523 ctggtctgttctcccaacacccaagaatgagacatctcaagcagactctatgtccagc 1582
 Db 1799 ctggttctccctcaacgaacataagatgactgttctgacatgaagctgtcgtcaggt 1858
 QY 1583 tggctcaagtgccaggggtcagagggaagacatcgataaggtgtccctatgctccctgct 1642
 Db 1859 ttgtcaagtgccataacttgaaggcgatgttcccaatgagcgcccttaccacgtgca 1518
 QY 1643 ctggaatccacacagcggtatctgcgttatagtggtcgtggagacttgaattgagcttc 1702
 Db 1919 gtgagtcacatgtgcgctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1978
 QY 1703 aaaaacatcacctgttggccaacaagtgttaccacaagaagtgaatgataatgcttcagt 1762
 Db 1979 gcaaacaccacatttcttgcacaataagtttgcacatgagatgtgaccccttgcctcag 2038
 QY 1763 gcttgaagaatcctacaggtatgaagccatccatcgtgagactggaacttggag 1813
 Db 2039 gtttgaagaacatctgagggacataaagcctgtgagacctaagaacgttaag 2089

RESULT 15
 AAC65468
 ID AAC65468 strand; cDNA; 5010 BP.
 XX
 AC AAC65468;
 XX
 DT 12-FEB-2001 (first entry)
 XX
 DE Diabetic rat heart core 2 GlcNAc-T coding sequence.
 XX
 KW Rat; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
 KW UDP-GlcNAc:Galbeta1-3GalNAc:alpha beta-1,6-N-acetylglucosaminyltransferase; ss.
 XX
 OS Rattus sp.
 XX
 PN US6131578-A.
 XX
 PD 17-OCT-2000.
 XX
 PF 02-OCT-1997; 97US-0943058.
 XX
 PR 02-OCT-1996; 96US-0046876.
 XX
 PA (KING/) KING G L.
 PA (NISH/) NISHIO Y.
 PA (KOYA/) KOYA D.
 PA (DENN/) DENNIS J W.
 PA (WARR/) WARREN C E.
 XX
 PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
 XX
 DR WPI: 2000-678642/66.
 DR P-PDB; AAB30297.
 XX
 PT Preventing or treating cardiomyopathy associated with diabetes mellitus
 PT and hyperglycaemia, comprises administering a substance that inhibits
 PT core 2 N-acetylglucosamine-T activity -
 XX
 PS Example 1; Fig 3A; 21pp; English.
 XX

CC The present invention describes the human UDP-GlcNAc:Galbeta1-3galNAc
CC alpha-hexosaminidase (core 2 GlcNAc-T) protein
CC and coding sequences. The enzyme is associated with cardiomyopathy in
CC diabetes and hyperglycemia sufferers. The invention also provides
CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
CC in treatment.

XX
SQ Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match 15.7%; Score 364.6; DB 21; Length 5010;
Best Local Similarity 58.3%; Pred. No. 1e-105;
Matches 659; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

Oy 686 caaagaaggtctatcacaggttcacccgaggggagccaaagagaggttcagg 745
Db 999 ccaataagaatgttaacttgcaccaaagtttcaagggtgacccagaagaatccgaagg 1018
Oy 746 ctatctgaataacctgtagtcaaa---gaagaagcgagagccttccacagacccact 802
Db 1039 tgaagcttgagatatacaagtcgaatcaagaagcgctccgagcgagacccgcatgact 1078
Oy 803 acctctccctcacagagagctgtgagcaactcaaggctgaaagaagtlcatagcttc 862
Db 1079 atataaacatgacccggactcgctccttcatcagacacgcaatatatatgagac 1138
Oy 863 cactgagcaagaagagtgtaggtccctattgcatctctatgtgtatctatgagaaga 922
Db 1139 cccttaccagaagaagaggtgtgcttccaatgtcatattccatagtggttcatcaataa 1198
Oy 923 ttgaaactttgaaaggtactctgagctgtgtatgcccctcagacatatatgtgtcc 982
Db 1199 ttgacatgttgacagagctccctgagggcatctatagccaagaatttctactgcatc 1258
Oy 983 atgttgatgagagagtcgccagaaattccaagaagcggtcaaaagaattattcttgc 1042
Db 1259 acgttgacagaaaaagacagagatccttlttagccggtgtagggcatgtgcatctgc 1318
Oy 1043 tcccaaatgtcttcaagccatgaagcttggtcggtgtgttatgctccctgtccaggg 1102
Db 1319 ttgataaatgtcttcttgccagccagcttggaagatgtgtatagcctccctcggagtc 1378
Oy 1103 tgaagcttgacactcaacttgtagaagacttgcacagagctcagtgccgttggaaatct 1162
Db 1379 ttaagctcgacactcaacttgtagaagagctgtacagaatgaaactggaactgact 1438
Oy 1163 tccgttaatacatgttgagcagagcttctctataaagaacatgcagaatgtgccaggtc 1222
Db 1439 tgaataatcttgtgtatgtgatttccctattaaacacacctggaattgtcaggaagc 1498
Oy 1223 tcaagaatttgatgaggaagtagacatgaggttcagaggttaacctctaagccaagaaga 1282
Db 1499 tcaagtcttccagagggaacacagctcgaaactggaagatgctccccaacaagaagag 1558
Oy 1283 ccggttggaatatcaactttaggttagtgaagacacatatacctaacaagaaga 1342
Db 1559 agaagtggaaaaaaacgatacaaggttggagcggaaagctgaacaacacttgagttgtca 1618
Oy 1343 aggatcctccctctataataattaactatgtttacaggaatgtcgtacatttgccttcc 1402
Db 1619 aggcgcagcctcactcaaaactcctctcttcttcagcgagcgtcatttgcgttacta 1678
Oy 1403 gagattcgttccaaatagtttttgaagaacccctaatacccaactgattgaatggttaa 1462
Db 1679 gggaaatagtggtatgtgtcgtgaataataaaatataatcaaaagtcatagatggtgac 1738
Oy 1463 aagaacattatagcccaatgacacccctctggccaaccttgaacgttgacggtgagtc 1522
Db 1739 aggaacataacagcccgataggttccctctggccaacatccaagaagatccctgaagtc 1798
Oy 1523 ctggtctgttcccaacaccccaagtlacgacatcagacatgacttctatgtccagtc 1582
Db 1799 ctggtctctccctcaagcctaagtagtactgtctgacatgaaatgtctgctgtaggt 1858

Oy 1583 tggtaagtgcagaggttcaatgaggagacatcgataagggtgtcctctatgtctcctgt 1642
Db 1859 ttgtcaatgagcaataactctgaaggcagatgttccaatggtgcgccttataccacgttga 1918
Oy 1643 ctgataccaccagcggtatctcggttataatgggcttgaggagacttgaattgactcttc 1702
Db 1919 gtgagatccatgtgtgctctgtgtgctctcttggagttgtgacttgaactgtgactgtgc 1978
Oy 1703 aaaccatcaactgttgcacaagaatttgaccacaaggtagaatgataatgtcttcaagt 1762
Db 1979 gaaacaccacttcttgccaataagttgacatgatgtgataccttgccttgcacag 2038
Oy 1763 gcttgaagaataactagttataaagccatctatgagactgaactttgag 1813
Db 2039 gtttggaagaacatctgagacataaagccctgtagagacttgaacgttaag 2089

Search completed: September 26, 2002, 08:19:19
Job time: 8458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:26:55 ; Search time 72.39 seconds
(without alignments)
7668.823 Million cell updates/sec

Title: US-09-874-390-1

Perfect score: 2319

Sequence: 1 ataactgggttcctcatl.....ataataacttaacaga 2319

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/CTOS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1990.8	85.8	2128	3	US-09-233-506-1
2	369.4	13.9	2105	1	US-07-955-041-3
3	369.4	15.9	2105	1	US-08-227-455-3
4	369.4	15.9	2105	1	US-08-472-482-3
5	369.4	15.9	2105	1	US-08-487-069-3
6	355	15.3	2102	3	US-09-063-237-3
7	180.4	7.8	1807	1	US-08-118-906-13
8	180.4	7.8	1807	1	US-08-486-196-13
9	180.4	7.8	1807	1	US-08-488-135-13
10	180.4	7.8	1807	2	US-08-474-065-13
11	162.8	7.0	378	1	US-08-118-906-3
12	162.8	7.0	378	1	US-08-486-196-3
13	162.8	7.0	378	1	US-08-488-135-3
14	162.8	7.0	378	2	US-08-474-065-3
15	134	5.8	378	1	US-08-118-906-1
16	134	5.8	378	1	US-08-486-196-1
17	134	5.8	378	1	US-08-488-135-1
18	134	5.8	378	2	US-08-474-065-1
19	129.6	5.6	192	3	US-09-233-506-9
20	84	3.6	147	3	US-09-233-506-13
21	43.6	1.9	66	1	US-08-118-906-11
22	43.6	1.9	66	1	US-08-486-196-11
23	43.6	1.9	66	1	US-08-488-135-11
24	43.6	1.9	66	2	US-08-474-065-11
25	43	1.9	99	1	US-08-118-906-7
26	43	1.9	99	1	US-08-486-196-7
27	43	1.9	99	1	US-08-488-135-7

28	43	1.9	99	2	US-08-474-065-7	Sequence 7, Appl1
c 29	35.4	1.5	7218	1	US-08-232-463-14	Sequence 14, Appl1
30	35	1.5	99	1	US-08-118-906-5	Sequence 5, Appl1
31	35	1.5	99	1	US-08-486-196-5	Sequence 5, Appl1
32	35	1.5	99	1	US-08-488-135-5	Sequence 5, Appl1
33	35	1.5	99	2	US-08-474-065-5	Sequence 5, Appl1
c 34	33.2	1.4	1941	4	US-09-402-002-1	Sequence 4, Appl1
c 35	33	1.4	11283	2	US-08-603-753D-3	Sequence 3, Appl1
c 36	33	1.4	11283	3	US-09-099-753-3	Sequence 3, Appl1
c 37	33	1.4	11283	4	US-08-986-106-3	Sequence 3, Appl1
38	32.8	1.4	246240	2	US-08-724-394A-20	Sequence 20, Appl1
39	32.8	1.4	246240	2	US-08-724-394A-21	Sequence 21, Appl1
40	32.8	1.4	246240	2	US-08-724-394A-22	Sequence 22, Appl1
41	32.2	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl1
42	32.2	1.4	3164	1	US-07-723-002C-3	Sequence 3, Appl1
43	32.2	1.4	5194	4	US-08-642-846-1	Sequence 1, Appl1
44	32.2	1.4	5194	4	US-09-264-604-1	Sequence 1, Appl1
45	31.8	1.4	900	1	US-08-218-026-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-233-506-1
Sequence 1, Application US/09233506
Patent No. 6,916,580

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
APPLICANT: Yeh, Jinn-Chern
TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
FILE REFERENCE: P-LI 3415
CURRENT APPLICATION NUMBER: US/09/233,506
CURRENT FILING DATE: 09/23/2001
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2128
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (354)..(1670)
US-09-233-506-1

Query Match 85.8%; Score 1990.8; DB 3; Length 2128;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2004; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY	244	agatataaagggaagcccgaaactgttcttgaacatttgaatgtcagaataac	303
DB	102	agatataaagggaagcccgaaactgttcttgaacatttgaatgtcagaataac	161
QY	304	tttggagggttagaagatcgaaggaatgtgttcacatttgcgcgaagaaacag	363
DB	162	tttggagggttagaagatcgaaggaatgtgttcacatttgcgcgaagaaacag	221
QY	364	ccagcttcaacttgaagaacgaatcagccttgaagaatcattccctaagaagag	423
DB	222	ccagcttcaacttgaagaacgaatcagccttgaagaatcattccctaagaagag	281
QY	424	aagctactaaagatgtgtctccctccacattcctgtgtcgtcctcactgtctcc	483
DB	282	aagctactaaagatgtgtctccctccacattcctgtgtcgtcctcactgtctcc	341
QY	484	cattctgtaacatgttcaatgaagagacctcgcagctgcattacttgggctcg	543
DB	342	cattctgtaacatgttcaatgaagagacctcgcagctgcattacttgggctcg	401
QY	544	ggctgcatatgcctggtgcacactgtgctctgaacttcttcagggttaagtgc	603
DB		ggctgcatatgcctggtgcacactgtgctctgaacttcttcagggttaagtgc	

Db 402 ggcgtctatatgctgctgccaactggtgctctgaaacttctcttcagggttgaaagtgtgac 461
Qy 604 tctgaccacttggtctggtgctcagggaatctcaaggccaactctgtagaatactctg 663
Db 462 tctgaccacttggtctggtgctcagggaatctcaaggccaactctgtagaatactctg 521
Qy 664 tataattctcgaacttccagcaagaaggtctatacaactgtgttcagggttccaggagg 723
Db 522 tataattctcgaacttccagcaagaaggtctatacaactgtgttcagggttccaggagg 581
Qy 724 gaccagaaggcaggtgttcagggtctatcttgataactgtgaggtcagaagaagcgagag 783
Db 582 gaccagaaggcaggtgttcagggtctatcttgataactgtgaggtcagaagaagcgagag 641
Qy 784 cctttccagacaccactcctcctccacagagactgtgagacttcaaggtcga 843
Db 642 cctttccagacaccactcctcctccacagagactgtgagacttcaaggtcga 701
Qy 844 aggaagttcatacagttcccactgagcaagaagggtggaagttccctatgtactct 903
Db 702 aggaagttcatacagttcccactgagcaagaagggtggaagttccctatgtactct 761
Qy 904 atggtgttatatgagaagatgtgaaactttgaaagctcagctcgtgtgtatgcccc 963
Db 762 atggtgttatatgagaagatgtgaaactttgaaagctcagctcgtgtgtatgcccc 821
Qy 964 cagaacataactgtgtcctcctgtagatgagaaggtccccaagaacttccaaaggcgctc 1023
Db 822 cagaacataactgtgtcctcctgtagatgagaaggtccccaagaacttccaaaggcgctc 881
Qy 1024 aaagcaattatttctgtctccccaagttctctcaagcaagtaagctgtgttcgggtgtc 1083
Db 882 aaagcaattatttctgtctccccaagttctctcaagcaagtaagctgtgttcgggtgtc 941
Qy 1084 tatgccccctggtcagaggtgcaagctgacccaactcagctgaggaagactgtctccagagc 1143
Db 942 tatgccccctggtcagaggtgcaagctgacccaactcagctgaggaagactgtctccagagc 1001
Qy 1144 tcaagtcggtggaatactctctgtaatacatgtggaagcgagacttccataaagagcaat 1203
Db 1002 tcaagtcggtggaatactctctgtaatacatgtggaagcgagacttccataaagagcaat 1061
Qy 1204 gcagagatgtgtccaggtctctcaagaatgttgatggaaggaataacatggaatcagaagta 1263
Db 1062 gcagagatgtgtccaggtctctcaagaatgttgatggaaggaataacatggaatcagaagta 1121
Qy 1264 cctcctaagcaagaagaacccgctggaataatcaactttgagtgagtgagagacacatta 1323
Db 1122 cctcctaagcaagaagaacccgctggaataatcaactttgagtgagtgagagacacatta 1181
Qy 1324 cacttaacccaagaagaagatcctccctataataatttaactatgtttacaggggaat 1383
Db 1182 cacttaacccaagaagaagatcctccctataataatttaactatgtttacaggggaat 1241
Qy 1384 gcgtacattgtggtctcccgagatttgcacaacatgtttggaagaacccctaattcccaa 1443
Db 1242 gcgtacattgtggtctcccgagatttgcacaacatgtttggaagaacccctaattcccaa 1301
Qy 1444 caactgtattgattgtaaaagacactatagccagatgacacccctctggtccaccctt 1503
Db 1302 caactgtattgattgtaaaagacactatagccagatgacacccctctggtccaccctt 1361
Qy 1504 cagcgtgacaggttgatgtctgtctgttcccaacacaccccaagtagcgaatctcagac 1563
Db 1362 cagcgtgacaggttgatgtctgtctgttcccaacacaccccaagtagcgaatctcagac 1421
Qy 1564 atgactctatgtccagagctgtgtcaagtgtcagaggtcatgtgaggaagacatcgaataaggt 1623
Db 1422 atgactctatgtccagagctgtgtcaagtgtcagaggtcatgtgaggaagacatcgaataaggt 1481
Qy 1624 gctcctatgtctcctcctcctggaatccacagcggtctatcctggtttatggggtcggg 1683
Db 1482 gctcctatgtctcctcctcctggaatccacagcggtctatcctggtttatggggtcggg 1541

Qy 1684 gacttgaattgagatgtcttcaaaaaccacactgtgtgccaacaagtttgacccaagagta 1743
Db 1542 gacttgaattgagatgtcttcaaaaaccacactgtgtgccaacaagtttgacccaagagta 1601
Qy 1744 gatgataatgtctctcagtggtcttgaagaagataactagttataagggcatctatgggact 1803
Db 1602 gatgataatgtctctcagtggtcttgaagaagataactagttataagggcatctatgggact 1661
Qy 1804 gaacttggagacacactatagaagcgttctgtaactgtgtgggccaagagcatgtacaacat 1863
Db 1662 gaacttggagacacacactatagaagcgttctgtaactgtgtgggccaagagcatgtacaacat 1721
Qy 1864 gctcagaactgtctgagacagtggtgggtgggagagccagggccttgcacatctgtgcatcc 1923
Db 1722 gctcagaactgtctgagacagtggtgggtgggagagccagggccttgcacatctgtgcatcc 1781
Qy 1924 tttagataaaggaggtcgtctattagatgtgtgggtaagtagatcttctgcttgcacattg 1983
Db 1782 tttagataaaggaggtcgtctattagatgtgtgggtaagtagatcttctgcttgcacattg 1841
Qy 1984 ctgcttggtggaatgctgctgtctcctcacccttaacccttagatgttctcctcaact 2043
Db 1842 ctgcttggtggaatgctgctgtctcctcacccttaacccttagatgttctcctcaact 1901
Qy 2044 ttctcactaagttagaattggaagactgtgtgataaggagaggttgaaggagataatgtgt 2103
Db 1902 ttctcactaagttagaattggaagactgtgtgataaggagaggttgaaggagataatgtgt 1961
Qy 2104 agagcacttgaattcaattgaattgacgtcgtgtagcttccacttctgttgaggtcgt 2163
Db 1962 agagcacttgaattcaattgaattgacgtcgtgtagcttccacttctgttgaggtcgt 2021
Qy 2164 tcccaataatccaggttgtgtagcgtggaaggagaacttgaattgaaagaaacacttccc 2223
Db 2022 tcccaataatccaggttgtgtagcgtggaaggagaacttgaattgaaagaaacacttccc 2079
Qy 2224 ttctgtactgttaacttaaaaaataa 2251
Db 2080 ttctgtactgttaacttaaaaaataa 2107

RESULT 2
US-07-955-041-3
; Sequence 3, Application US/07955041
; Patent No. 5360733
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; TITLE OF INVENTION: A NOVEL BETAL-6
; TITLE OF INVENTION: N-ACTIVATEDCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSILIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMATIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-POS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,041
; FILING DATE: 19921001
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815

```

REFERENCE/DOCKET NUMBER: P-LJ 9294
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2105 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..1504
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1913..1918
FEATURE:
NAME/KEY: misc_signal
LOCATION: 248..314
OTHER INFORMATION: /standard.name=
OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-07-955-041-3

```

```

Query Match          15.9%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. 9.4e-109;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

```

OY 772 aagaagcagagccttccacagacacccactcctccctccacagagctgtgacac 831
DB 466 AAAAAAGCCCGCGGTGGACCTGACGACTATATAACATGACGAGTGGTCTTCT 525
OY 832 ttcaagcgtgaagaagatctacagttccctcagcagaagaagaggttgatccct 891
DB 526 TTCATCAAGAGAGCGCAATATATTTAGAACCCCTTAGTAAGAAGAGGCGGATTTCCA 585
OY 892 atgcatactatgtatcatgagaagattgaaacttgaaagcttactgcagact 951
DB 586 ATAGCATATTTCTATAGTGTTCATCAAGATTGAATCTTGACAGGCTGCTGAGGCGC 645
OY 952 gtctatgccccctagaacatactgtgtcatalgtgagatgaaagtcgccagaacttc 1011
DB 646 ATCTATATGCTCCAGCAAAATTTCTATTGCGTTTCATGTGACACAAATCCGAGATTCCTAT 705
OY 1012 aaagagcggtcacaagaatattcttctgtctcccaatgtcttccacgttaagctg 1071
DB 706 TTAGCTGAGTGTATGAGGATCGCTTCTGTTTAAATGTTCTTGTGCGCCGCGATTG 765
OY 1072 gtccggtgttatagtcctcctcgtcccaaggtgagcagctgacccaactgagaaagc 1131
DB 766 GAGAGTGTGTTTATGTCATCGTGAGCCGGGTTCAAGGCTGACCTCACTGCAAGAGAT 825
OY 1132 ttgtccagagctcagtcgcgtggaatactcctcgaatatactgtggaacgagacttcct 1191
DB 826 CTCTATGCAATGAGTGAACCTGGAAGTACTGTATTAATCTTTGATGATGATTTTCCC 885
OY 1192 ataagaagcaatgcaagaatgtgtccaggtctcgaagatgttgaayggaagagatgcatg 1251
DB 886 ATTAAACCAACTGAATATGTTGACGAAGCTCAAGTTGTTAATGGGAAAAACAACCTG 945
OY 1252 gagtcaaggttaccctcgaagacaaagaaacccgctgaatatcactttgagtagtg 1311
DB 946 GAAACGAGAGAGATGTCATCCCATTAAGAGAAAGGTGAGAGACGGATATAGGTCGTT 1005
OY 1312 agagacacattacacccaacacaagaagaagatcctccctcctataatlaactatg 1371
DB 1006 AATGGAAGGCTGACAAACACAGGAGCTGTCAAAATGCTTCCCACTGCAACACCTCTC 1065
OY 1372 tttaaggaatagcgtaacatgtgtgtctcccgagatltcgtccacaatgttttgaagaac 1431
DB 1066 TTTTGTGCGAGTGCCTACTCTGTGTGACAGTAGGAGATGTGGGGTATGTACTACAGAAAT 1125

```

```

OY 1432 cctaaatcccaacaactgtatgtatgtaanaagacacttatagccacagatgaacactc 1491
DB 1126 GAAAAATTCAAAAGTGTATGAGTGGGCAACAAGACATATACAGCCCTGATGATATCTC 1185
OY 1492 tgggcccaccctcagcgtgtcaaggtgtgagtcgtgtcgttcccaaccccccaatgc 1551
DB 1186 TGGGCCACCATCCAAAGGATTCCTGAAGTCCCGGGCTCACTCCTCCAGCATTAAGTAT 1245
OY 1552 gacatctcagacatgactctatctccagagcgtgtcgaatgtgcaaggtcattaggaagac 1611
DB 1246 GATCTATCTGACATGACAGACAGTGGCAGTTGTCAAGTGGACAGTATTTGAGGGTGAN 1305
OY 1612 atcgataaggtgtcctcctatgtcctcgtcgtcgtcgtcgaatccacagcgggctatcgtt 1671
DB 1306 GTTTCAGAGGCTGCTCCCTACCCGCCCTCGATGAGTGCATGTGCGCTCAGTGTGATTT 1365
OY 1672 tatgggctgggacttgatgtgagttcctcaaacatcactgtgtgccaacaagttt 1731
DB 1366 TTCGGAGCTGTGACTTGAATGTGATGTCGCCCAACACCACTTGTTCCTCAATAAGTTT 1425
OY 1732 gaccacaagtagatagatagtcctcctcagtgctttagaagaataactagttataagc 1790
DB 1426 GACGTGAGATGTGACCTCTTTCACATCCAGTGTGTGATGAGCATTTGAGACACAAAGC 1484

```

```

RESULT 3
US-08-227-455-3
Sequence 3, Application US/08227455
Patent No. 5624832

```

```

GENERAL INFORMATION:
APPLICANT: FUKUDA, MINORU
APPLICANT: BIERHUIZEN, MARTI FA
TITLE OF INVENTION: A NOVEL BETA-5
TITLE OF INVENTION: N-ACETYLGLUCOSAMINITRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,455
FILING DATE: 14-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHERYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9957
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2105 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..1504
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1913..1918

```

```

FEATURE:
NAME/KEY: misc-signal
LOCATION: 248..314
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-227-455-3

```

```

Query Match      15.9%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 9.4e-109;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

```

QY 772 aagaagcagagccttcacagacaccactcctccctcaccagagactgtgagcac 831
DB 466 AAAAAGCGCCCTCGTGACACCTGACGACTATATTAACATGACAGTGTCTTCTCT 525
QY 832 ttaagagctgaaggaaggttcatacagttccctcctgagcaagaagggagggtccct 891
DB 526 TTCAATCAGAGACGCAATATATTTGTAAGAACCCCTTAGTAAGAGAGGCGAGTTTCCA 585
QY 892 attgacatactatgtgtacatgagagatigaanaacttgaagaagctactcgaagct 951
DB 586 ATAGCATATTTCTATAGTGGTTATCACAAGATTGAATGCTTGACAGGCTGCTGAGGGCC 645
QY 952 gtgtatgcccctcagaacatactgtgtccatgtgagatgaagaagttcccaagaacttc 1011
DB 646 ATCTATATGCTCAGATTTCTATTTGCTTCATGTGACACCAAAATCCGAGATTTCTAT 705
QY 1012 aagaagcagcgtcacaagaataattcttctcctcccaaatctcctcaccagagactgt 1071
DB 706 TTAGCTCAGATGATGGGCAATCGCTTCTGTTTATGTAATGTTCTTTTGCCGACCGCATTTG 765
QY 1072 gtctgggtgtgttatagctcctgtgtccagagtgccaagctacactcaactcatgtgaagac 1131
DB 766 GAGAGTGTGGTTATGATGCTGTGAGCGGGTTCAGGCTGACCTCACTGATGATGAAGAT 825
QY 1132 ttgtccagagctcagtgccgttggaatactcctgaatacatgtggagcaggtctcct 1191
DB 826 CTCTATGCAATGAGTGCAAACTGCAAGTACTGTATTAATCTTTGTGATGATGATTTTCCC 885
QY 1192 ataaagcagatcagagatggtccagagctcctcagaatgttgaaatggaggaatagatg 1251
DB 886 ATTAAACCAACTAGAAATTTGTCAGAGACTCAAGTTTGAATGGAGAAACAAACCTG 945
QY 1252 gaagtcagaggttacctcctcacaagaacacccgctcgtgaataatcactttagagtagtg 1311
DB 946 GAAGCGAGAGGATGCAATCCCAATAAGAAAGAAAGTGAAGAAAGCGTATGAGTGTGTT 1005
QY 1312 agagacacattacaactacaacaagaagaagatcctcccttaatttaactatg 1371
DB 1006 AATGGAAAGCTGACAAACACAGGAGACTGCAAAATGCTTCTCCACTCGAAMACACTCTC 1065
QY 1372 tttaacaggaatcgatcatctgtgtcctccagagattcgttccaacatgttttgaagaac 1431
DB 1066 TTTTCTG6CAGTGCCTTCTGCTGCTAGTAGGAGATATGTGGGTATGTACTACAGAAAT 1125
QY 1432 cctaaatcccaacaactatgtatgggtaaaagacactatagccccagaatgaacactc 1491
DB 1126 GAAAAAATCCAAAAGTTGATGAGAGTGGGCAAGACACATACAGCCCTGATGATATCTC 1185
QY 1492 tgggccaaccttcaagctgtcagcggtgagatgcctgtgctgttcccaacaccccaagtaac 1551
DB 1186 TGGGCGCACCTCCAAAGAGATTTCTGAAAGTCCCGGCTCACTCCCTGCACACCAATATAT 1245
QY 1552 gaacatcacaacatgaacttatttccagagctgttcaagtgagcagggatcaltgagggagac 1611
DB 1246 GATCTATCTACATGACAGAGGTTGCTCAAGTGTGCAAGTGGCAGTACTTTGAGAGGTGAT 1305
QY 1612 atcgataaggtgtcctcctatgtcctcctgtcttgaataccacagcgggctatcgtt 1671
DB 1306 GTTTCAGAGGTGCTGCTCCCTACCGCCCTCGGATGAGATGCATGTGCGCTCAGTGTGATAT 1365
QY 1672 tatggggcttgggagacttgaattgagatgttcaaaaacatcaactgttggccaacaagttc 1731

```

```

DB 1366 TTCGAGACTGCTGACTGCACTGCACTGCTCGCAAAACCACTGTTGCAATTAAGTTT 1425
QY 1732 gaccacaagtagatgtaagtctcctcagtgctgtcgttgaagaataaccagctttaaagc 1790
DB 1426 GACGTGAGATGTGACCTCTTTGCAATCAGTGTGATGAGCATTTTGAACACAAAGC 1484

```

```

RESULT 4
US-08-472-482-3
Sequence 3, Application US/08472482
Patent No. 5658778

```

```

GENERAL INFORMATION:
APPLICANT: FUKUDA, MINORU
APPLICANT: BIERHUIZEN, MARTI FA
TITLE OF INVENTION: A NOVEL BETAI-6
TITLE OF INVENTION: N-ACETYLGLYCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUCOSTALIN AND A METHOD FOR CLONING PROTEINS HAVING
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,482
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/955,041
FILING DATE: 01-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9294
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2105 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..1504
FEATURE:
NAME/KEY: polyA-signal
LOCATION: 1913..1918
FEATURE:
NAME/KEY: misc-signal
LOCATION: 248..314
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-472-482-3

```

```

Query Match      15.9%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 9.4e-109;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

```

QY 772 aagaagcagagccttcacagacaccactcctccctcaccagagactgtgagcac 831
DB 466 AAAAAGCGCCCTCGTGACACCTGACGACTATATTAACATGACAGTGTCTTCTCT 525

```



```

Db 466 AAAAAAGCCCTCGGTGGACACCTGACGACATATATAACATGACCAGTACTTCTTCT 525
Oy 882 ttcaagctgaaagaagttaacagttccacagcaagaagagtgagttccct 891
Db 536 TTCAATCAAGAGAGCGCAAAATATTTGTAGAACCCCTTAGTAAAGAGAGCGGATTTTCCA 585
Oy 892 attgatactactgtatgtatcagaagaattgaagaacttgaagaagctactgcgagct 951
Db 586 ATAGCATATTTCTATAGTGGTTCATCAACAAGATTGAAGCTTGACAGAGGTGCTGAGGGCC 645
Oy 952 gtgtatgccctcagaacatactgtgtccatgtgatgaaagtcgccagaacttc 1011
Db 646 ATCTATATGCGCTCAAGAAATTTCTATTGCGTTCATGTGGACACAAATAATCCGAGGATTCCTAT 705
Oy 1012 aaagagcggtcaagaagaactaattctgtctcccaaatgtctcatagcgagtaagctg 1071
Db 706 TTAGCTGCAAGTATGGATGCGCTTCCTGTTTAACTATAGTCTTGTGTGGCAGCCGATTTG 765
Oy 1072 gtccggtgttattatgtcctcctgtgccagagtgcaagctcaactcaactgtcaagaag 1131
Db 766 GAGAGTGTGGTTTATGCAATCGTGAGCGGGTTCAAGCGTCAACTGATGATGAAGAT 825
Oy 1132 ttgtccagagctcaagtcgcgttgaacactcctgaatacactgtggagcagacttcct 1191
Db 826 CTCTATGCAATGAGTGCACAACTGGAAGTACTTGATAAATCTTTGTGATGATTTTCCC 885
Oy 1192 ataaagcaatgacagaagaatgtccaggtctcagaatgttgaatggagagatgacatg 1251
Db 886 ATTTAAACCAACCTTAGAAATTTGTGAGAAAGCTCAAGTTTATGAGGAGAAACAAACCTG 945
Oy 1252 gactcagagtgtaactcctcaagaacacacacccgctggaatcactcattgagtgatg 1311
Db 946 GAAACGGAGAGAGATCCCATTAAGAAGAAAGTGTGAACAAACGGATATAGGTCGTT 1005
Oy 1312 agagacacattacactcaacaaagaagaagatcctccctataatlaactatg 1371
Db 1006 AATGGAAGCTGCACAAACAGAGGACTGTCAAAATGCTTCCTCCACCTGAAACACCTCTTC 1065
Oy 1372 ttacaggaatgtgtacatgtgtgtcccgagatttcgcaacatgttttgaagaac 1431
Db 1066 TTTTCTGCGACGTCTACTCTGTGTCAGTATGAGGATATGTGGGTATGTACTTACAGAT 1125
Oy 1432 cctaaatcccaacaactgattgaatggglaaagaacacttataagccagatgaacactc 1491
Db 1126 GAAAAAATCCAAAGTTGATGAGTGGCACAACACATACAGCCCTGATAGTATCTC 1185
Oy 1492 tgggcacaccttcagcgtgcgaagtgatgtcgtgtctgttcccaaccccccaagtac 1551
Db 1186 TGGGCGCACCATCCAAAGATTCCTGAAGTCCCGGGCTCACTCCCGCAGCCCATAGATAT 1245
Oy 1552 gacactcagacatgacttctcattgcacaggtcgtcgaatgtgcaaggtgcaatgaaggagac 1611
Db 1246 GATCTATCTGACATCCAAAGCAGTTGCCAGTTTGTCAAGTGGCACTTCTTAGGGGTGAT 1305
Oy 1612 atcgaataaggtgtcctcattatgtcctcgtctgtgaatccacagcggtcactcgtcgt 1671
Db 1306 GTTTCGAAGGGGTCTCCCTACCCCGCTGCGATGAGTCCATGTCGCGCATGTGCAAT 1365
Oy 1672 tatggagctggaggaactgaattgtgatgtctcaaaacacactcgttggcacaagttt 1731
Db 1366 TTGGAGGTGTGACTTTTAACCTGATGCTGGCAACACCATCTTGTGCCAATTAAGTTT 1425
Oy 1732 gaccacaaggtatgataatgtcctcattagatgtctagaagaatcactcgttataagc 1790
Db 1426 GACGTGATGTTGACCTCTTTCCTCCATCCAGTGTGATGACATTTTGAACACAAAGC 1484

```

```

RESULT 5
US-08-487-069-3
: Sequence 3, Application us/08487069
: Patent No. 5684134
: GENERAL INFORMATION:
: APPLICANT: FUKUDA, MINORU

```

```

: APPLICANT: BIERHUIZEN, MARTI FA
: TITLE OF INVENTION: A NOVEL BETA1-6
: TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
: TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
: TITLE OF INVENTION: ENZYMATIC ACTIVITY
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CAMPBELL AND FLORES
: STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
: CITY: SAN DIEGO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: us/08/487,069
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/955,041
: FILING DATE: 01-OCT-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: CAMPBELL, CATHRYN
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9294
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-535-9001
: TELEFAX: 619-535-8949
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2105 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 220..1504
: FEATURE:
: NAME/KEY: POLYA_signal
: LOCATION: 1913..1918
: FEATURE:
: NAME/KEY: misc_signal
: LOCATION: 248..314
: OTHER INFORMATION: /standard_name=
: OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
: US-08-487-069-3

```

```

Query Match 15.9%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 9,4e-109;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

```

Oy 772 aagaagcgagagccttccacagaacaccactcctcctcaccagaagactgtgaagac 831
Db 466 AAAAAAGCCCTCGGTGGACACCTGACGACATATATAACATGACCAGTACTTCTTCT 525
Oy 832 ttcaagctgaaagaagttaacagttccacagcaagaagagtgagttccct 891
Db 526 TTCAATCAAGAGAGCGCAAAATATTTGTAGAACCCCTTAGTAAAGAGAGCGGATTTTCCA 585
Oy 892 attgatactactgtatgtatcagaagaattgaagaacttgaagaagctactgcgagct 951
Db 586 ATAGCATATTTCTATAGTGGTTCATCAACAAGATTGAAGCTTGACAGAGGTGCTGAGGGCC 645
Oy 952 gtgtatgccctcagaacatactgtgtccatgtgatgaaagtcgccagaacttc 1011
Db 646 ATCTATATGCGCTCAAGAAATTTCTATTGCGTTCATGTGGACACAAATAATCCGAGGATTCCTAT 705

```

```

QY 1012 aaagagcggtcaagaacatattcttgcctcccaaatctctcaccagtaagctg 1071
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 TTGAGTGGACATGATGGGCAATCGCTCTGTTTGTATGATCTTTGTGGCCGATG 765
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1072 gtccgggtgttatagctcctgtgccagggtgcaagctaccacacatgagagac 1131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 GAGAGTGTGTTTATGATGCTGTGGAGCCGGTTTCAGGCTGACCTCACTGATGAAGAT 825
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1132 ttgctccagagctcagtgccgttggaatctctcgaatacaatgtggagagcttct 1191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 CTCTATGCAATGATGCAAACTGGAAGTACTGTATTAATCTTTGTGTATGATTTTCCC 885
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1192 ataaagagcaatgcaagatggtccaggctctcaagatgttgtaatggaggaatagcatg 1251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 ATTAAACCAACCTAGAAATGTTGAGAACCTCAAGTTGTTAAAGGAGAAACAACTG 945
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1252 gattcaagaagttacctcctaagcaaaagaaacccgctggaatctcacttgagtggtg 1311
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 946 GAAACGAGAGATGATGCTATCCCATAAAGAAAGAGTGAAGAGCGGTATGAGGTGTT 1005
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1312 agagacacattacacctcaacaagaagaagatcctccccccttaatttaactatg 1371
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1006 AATGGAAGAGTGAACAACAGAGGACTGTCAAAATGCTTCTCCACACGGAACACTCTC 1065
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1372 tttaacaggaatgctacattgtgtgtctcccgagattcgtlccaacatgtttgaagaac 1431
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1066 TTTTCTGGGAGTGGCTACTTCTGTCAGTATGAGGAGTATGGGGGTATGATACAGAAAT 1125
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1432 cctaatacccaacaacagcttgtaatggtaaaagacactatagccagatgaaacactc 1491
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1126 GAAAAAATCCAAAAGTTGATGAGTGGGCAACAGACATACACGCTGTGATGATATCTC 1185
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1492 tgggccaacccctcagctgacagtgatgagatgctgtctgttcccaaccccccaagtg 1551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1186 TGGGCCACATCCAAAGATATCTCGAAGTCCCGGCTCAGTCCCTGCCAGCATAAATAT 1245
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1552 gacatctcaacatgactcttatgtccaggctgtgcaagtggcagggtlcatgaggaagac 1611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1246 GATCTATCTACATGCAAGAGATGCAAGGTTTGTCAAGTGCGACTTTTGAAGGTGAT 1305
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1612 atcgataaggtgctcctctatgtcctcgtctgtgaaatccacagcgaggtatctgctt 1671
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1306 GTTTCGAAGGCTGCTCCCTTACCCGCTTGGAGTGAAGTCCATGTGCGCTCATGTGCAAT 1365
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1672 tatgggctgggagactgaaatgtgattgtctcacaacacatcactgtgtggcacaagaatt 1731
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1366 TTGCGAGCTGTGACTGTGAAGTGTGATGCTGCCCAACACACACTGTGTTGCCAATAGTTT 1425
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1732 gaccacaagttagatgataatgtctctcactcagtgcttagaagaatacctcagttataaggc 1790
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1426 GACGTGATGTGACCTCTTTGCGCATCCAGTGTGATGAGCAATTTGAGACACCAAGC 1484
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6
US-09-063-237-3

```

; Sequence 3, Application US/09063237
; Patent No. 6124267
; GENERAL INFORMATION:
; APPLICANT: McEver, Roger P.
; APPLICANT: Cummings, Richard P.
; TITLE OF INVENTION: O-Glycan Inhibitors of Selectin Mediated
; TITLE OF INVENTION: Inflammation Derived from P5G1-1
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree
; STREET: Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30306-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,237
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/649,802
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRP110CIP7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)873-8794
; TELEFAX: (404)873-8795
; INFORMATION FOR SEQ. ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-063-237-3

Query Match 15.3%; Score 355; DB 3; Length 2102;
Best Local Similarity 59.3%; Pred. No. 4,2e-104;
Matches 604; Conservative 0; Mismatches 415; Indels 0; Gaps 0;

QY 772 aagaagcgagagccttcacagacaccactacctccctcaccagagaactgtgagac 831
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 AAAAAAGCGCCGCTGGTGAACCTGACGACTATATAAACAATGACAGTGTCTTCT 524
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 832 ttcaaggctgaaggaagttcatagagttccactgagcaagaaggggtgagttccct 891
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 TTGATCAAGAGACGCAATATATTGTGAACCCCTTGTAAAGAGGGCGGAGTTTCA 584
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 892 attgacactctatgtgattctgaaagattgaaacattgaaagctactgagagct 951
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 ATGACATATCTATAGTGTGTTCAACACAGATTAATGCTTGACAGCTGTGAGGCC 644
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 952 gtgtatgcccccaagaacatatactgtgtcatgtggaagagagtcaccagaacttc 1011
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 ATCTATATGCTCAGATTTCTATTGCTTCAATGTGACACAAATCCGAGATTCCTAT 704
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1012 aaagagcggtcaagaacatattcttgcctcccaatgtctcactagcagtaagctg 1071
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 705 TTAGCTGACGTGATGGGCAATGCTTCTGTTTATGATATGTTGTGGCCAGCCGATG 764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1072 gtccgggtgttatagctcctgtgccagggtgcaagctgacactcaactgcatggaagac 1131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 765 GAGAGTGTGTTTATGATGCTGTGGAGCCGGTTTCAGGCTGACCTCACTGATGAAGAT 824
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1132 ttgctccagagctcagtgccgttggaatctctcgaatacaatgtggagagcttct 1191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 825 CTCTATGCAATGATGCAAACTGGAAGTACTGTATTAATCTTTGTGTATGATTTTCCC 884
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1192 ataaagagcaatgcaagatggtccaggctctcaagatgttgtaatggaggaatagcatg 1251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 885 ATTAAACCAACCTAGAAATGTTGAGAACCTCAAGTTGTTAAAGGAGAAACAACTG 944
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1252 gattcaagaagttacctcctaagcaaaagaaacccgctggaatctcacttgagtggtg 1311
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 945 GAAACGAGAGATGATGCTATCCCATAAAGAAAGAGTGAAGAGCGGTATGAGGTGTT 1004
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1312 agagacacattacacctcaacaagaagaagatcctccccccttaatttaactatg 1371
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1005 AATGGAAGAGTGAACAACAGAGGACTGTCAAAATGCTTCTCCACACGGAACACTCTC 1064
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1372 tttaacaggaatgctacattgtgtgtctcccgagattcgtlccaacatgtttgaagaac 1431
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1065 TTCTGCGAGTGCCTACTTCGTCCTCACTAGGAGCATATGTGGGTATGTACTACAGAAAT 1124
QY 1432 cctaaatcccaactgaactgaatggtgaaagaacactatagcccgatgaacaccc 1491
Db 1125 GAAAAATCCAAAGTTGAGTAGTGCGCAAGACACATACAGCCCTATGATGATCTTC 1184
QY 1492 tgggcacccctcagcgtgcaagtgagatgctgctgctcccaaccaccccaagtc 1551
Db 1185 TCGCGCACATCCAAAGATTCCTGAAGTCGCCGCTCACTCCCTCCAGCCATTAAGTAT 1244
QY 1552 gacatccagacatgaactctatctgccaagcctgltcaagtggaaggtcaatgaaggagac 1611
Db 1245 GATCTATGTGATGCAACAGTCCAGCTTGTCAAGTGCAGCACTTTCAGGAGTAT 1304
QY 1612 atcgataaagtggtcctcctatgctccctgctctggaatccaccaggggtctatcgctt 1671
Db 1305 GTTTCCAAAGGTCCTCCCTACCCGCTCGATGAGTCCATGTCAGTGGCTCACTGTGCATTT 1364
QY 1672 tatggagcctggggaacttgatgagatgcttcaaaacacacactgttggccaagaattt 1731
Db 1365 TTCGGAGCTGCTGACTGTGACATGAGATGCTGCCCAAAACACACTTGTTCCTCAATACCTT 1424
QY 1732 gacccaagtgatgataatgctctcctcaagtgtccttagaagaatccctacgttataaagc 1790
Db 1425 GACGTGCATGTGACCTCTTTCCTCCATCCAGTGTTCATGACATTTGAGACAAAGC 1483

```

RESULT 7

US-08-118-906-13

; Sequence 13, Application US/08118906

; Patent No. 5484590

; GENERAL INFORMATION:

; APPLICANT: Fukuda, Minoru

; TITLE OF INVENTION: Expression of the Developmental I

; TITLE OF INVENTION: Antigen By a Cloned Human CDNA Encoding a Member of a

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/118,906

; FILING DATE: 09-SEP-1993

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1807 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 255..1454

; US-08-118-906-13

Query Match 7.8%; Score 180.4; DB 1; Length 1807;
 Best Local Similarity 54.3%; Pred. No. 8,5e-48;
 Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

```

QY 823 tgtgacacttcaagagctgaagaagttacacagttccaccctgagcaagaagagtg 882
Db 468 TCGAAGAAATTAAGTCCAGAGCCACCTATACAGAGCCCTTTATTAAGGAAGAGCT 527
QY 883 gaattccctatgtacactctatggtgatcctatggaagaattgaaacttgaagacta 942
Db 528 GACTTTCCTTGGCATATTAATGATGATCAATCACTTTGACACCTTTCAGAGGCTC 587
QY 943 ctggagcgtgtatgacccctcagaacataactctgttccatggtatgaatgagcccca 1002
Db 588 TTCAGGCTATTTACATGTCGCCAATAATATCTACTGTGTTCAATGTGATGATGAAGACACA 647
QY 1003 gaacttccaagagcggtlcaaaagcaatlatcttctgcttcccaatgtctctatagcc 1062
Db 648 ACTGAATTTAAAGATGCGGTAGAGCAATTAAGCTGCTTCCAAAGCCTTTTGCGCT 707
QY 1063 agtaagctggttcgggtgttatgctctcctcgttccaggtgcaagctgaactcaatgc 1122
Db 708 TCCAAAGATGGAACCCGTTCTATGAGGAGATCTCCAGGCTCCAGGCTGACCTGAAGCTGC 767
QY 1123 atggaagacttgctccagagctcagtgccgttggaatactctgatacatggtgagacg 1182
Db 768 ATCAGAGATCTTTCGCTTCAGAGTCTATGAGATGAGTATCAACACCTGTGGGCA 827
QY 1183 gacttctctaataagaagcaatgacagatggtccaggtctcagaagtgttgaatggag 1242
Db 828 GACTTCCCTCCGTAAGAACCAAGAAATAGTTAGTATGTAAGATTTAAAGATAA 887
QY 1243 aatgcatgagtgacagaggttacctcctaagcaaaagaaccgctggaataat----- 1296
Db 888 AATATATCCCAAGGGGTGTCGCCCAAGCTCATGCAATTGACGAGCAATTAATGTCCAC 947
QY 1297 -----cacttggagtgatgagagacacattaccactaccacaagaagaagatctc 1350
Db 948 CAAAGACACTGGGCAAGAACGCTTCTCTATGATTAAGAACCAAGACGTTGAACCGCT 1007
QY 1351 cccctataacttaactatgtttacaaggaatgcgtacatgtgctcccgagattc 1410
Db 1008 CCCCCCATTAATCTCACAATTTACTTTGGCTCGCTTATGTGCTATCAAGAGAGTTT 1067
QY 1411 gtccaaatggtttgaagaaccttaattcccaacactgattgaatgggttaagaacct 1470
Db 1068 GCCAACTTTGTTCTGCATGACCCACGCGCTGTTGATTTGCTCCAGTGTCAGAGCACT 1127
QY 1471 tatagccagatgaacactctgggccacccttcaacgctgcaagtggtgctgctct 1530
Db 1128 TTCAGTCCCTGATGAGCATTTCTGGGTGACATCAATAGATTCAGAGTCTTCTGCTCT 1187
QY 1531 gtcccca 1538
Db 1188 ATGCCAAA 1195

```

RESULT 8

US-08-486-196-13

; Sequence 13, Application US/08486196

; Patent No. 5731420

; GENERAL INFORMATION:

; APPLICANT: Fukuda, Minoru

; TITLE OF INVENTION: Expression of the Developmental I

; TITLE OF INVENTION: Antigen By a Cloned Human CDNA Encoding a Member of a

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; US-08-486-196-13

```

STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,196
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1807 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 255..1454
US-08-486-196-13

```

```

Query Match          7.8%; Score 180.4; DB 1; Length 1807;
Best Local Similarity 54.3%; Pred. No. 8.5e-48;
Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

QY 823 tgttagcactcgaagctgtaaggaagttcatcacgtccacgtgagaagaagagtg 882
    || || || || || || || || || || || || || || || || || || || ||
DB 468 TGCAGAGAAATCTGACCCAGAGCAGACATACAGCCCTTTATCAAGAGAAGAGCT 527

QY 883 gacttccctatgcatactctatgtgtattcattgagaagattgaaacttgaagagcta 942
    || || || || || || || || || || || || || || || || || || || ||
DB 528 GACCTTCCCTTGCGATATATATAGTCAATCATCATCTTGACACCTTGCAAGGCTC 587

QY 943 ctgcgagctgtgtatgccccctcagaacatatactgtgtccatgtgagatgaaagtc 1002
    || || || || || || || || || || || || || || || || || || || ||
DB 588 TTCAGGGCTATTATGATGCCCAAAATATCTACTGTCTTCTCATGTGATGAAAAAGCAACA 647

QY 1003 gaaacttcaagaagcggtlcaaaagaattatcttctgtctcccaatgtctcatagcc 1062
    || || || || || || || || || || || || || || || || || || || ||
DB 648 ACTGAATTTAAAGATGCGGTAGACACTATTAGCTGCTCCCAAGCGTTTCTGTGCT 707

QY 1063 agtaagctgtgtgtgtgtgtatgtatgtctccgtgtccagagtgatgaagctgacctaa 1122
    || || || || || || || || || || || || || || || || || || || ||
DB 708 TCCAAGATGGAACCCGTGTCTATGAGGGATCTCAAGGCTCCAGGCTGACCTGAACCTGC 767

QY 1123 atggaagaactgtgtccagagtgccgtgtggaataactctccgaatactatgtggagcg 1182
    || || || || || || || || || || || || || || || || || || || ||
DB 768 ATCAGAGATCTTTCTGCTTGCCTTGAGGTCTCATGAGATGTTATTCACACCTGTGGGCAA 827

QY 1183 gacttccctataaagaagcaatgagatgtlccaggctcctcaagatgttgaatggagag 1242
    || || || || || || || || || || || || || || || || || || || ||
DB 828 GACTTCCCTCGAAGAACCAACAAGAAATAGTTCAGTATTCGAAGAAGTTAAAGGTAAA 887

QY 1243 aatagcatgtgagtgagagtgactcctctaagcaagaagaaccgctggaaatat----- 1296
    || || || || || || || || || || || || || || || || || || || ||
DB 888 AATATGACCCCGAGGGGTGCTGCCGCCAGCTCATGCAATTTGGACGAGACTAATATATGTCAC 947

QY 1297 -----cacttggaggtgagtgagagacaattacaccttaaccacaagaagaagatcct 1350
    || || || || || || || || || || || || || || || || || || || ||

```

```

DB 948 CAAAGAGACCTGGCAAAGAGCTTTCTATGTGATAAGAACAAACAGCGTTGAAACCGCCT 1007
QY 1351 cccctataattaactatgtttacaggaatgctgacattgtgtctccagatttc 1410
    || || || || || || || || || || || || || || || || || || || ||
DB 1008 CCCCCCATATATCTCACAAATTTACTTGTGGCTCTGCTCATGTGCTCATCAAGAGATTT 1067

QY 1411 gtccaaatgttttgaagaaccctaataatccaaacattgattgaatggagtaaaagacct 1470
    || || || || || || || || || || || || || || || || || || || ||
DB 1068 GCCAACTTGTGTCGATGACCCACGCGGCTGTTGATTTGCTCCAGTGTCCAAAGACACT 1127

QY 1471 tatagcccaatgaaacacccctgtggccaccccttcagcggtgagagtgatgctgtct 1530
    || || || || || || || || || || || || || || || || || || || ||
DB 1128 TTCAGTCTGATGAGCAATTTCTGTGGGTGACACTCAATAGATTCCAGGTGTTCCTGGCTCT 1187

QY 1531 gtcccaa 1538
    || || ||
DB 1188 ATGCCAAA 1195

```

```

RESULT 9
US-08-488-135-13
Sequence 13, Application US/08488135
Patent No. 5766910
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,135
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1807 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 255..1454
US-08-488-135-13

```

```

Query Match          7.8%; Score 180.4; DB 1; Length 1807;
Best Local Similarity 54.3%; Pred. No. 8.5e-48;
Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

```

```

Oy 823 tgtgagcacttcaagcctgaaagaagltcacatagcttccacgtgagcaagaagtg 882
Db 468 TGCAAGAAATCTTGACCCAGAGCCACTACATCAAGGCCCTTTATCTAAGGAAGAAGCT 527
Oy 883 gagttccctatgtcactctatcgttgatcagtgagaagatgaaacttgaaagcta 942
Db 528 GACTTTCCTTGCGATATATATATAGGTATCATCATCTTGTGACACCTTTGCAAGGCTC 587
Oy 943 ctgcagactgtgtatgcccctcaagacatatagctgtccatgtgtgagagaagcccca 1002
Db 568 TTCAGGGCTATTTACATGCCCCCAAAATATCTACTGTCTTCATGTGATGAAAAAGCAACA 647
Oy 1003 gaaacttcaagaagcggtgcaagaacatattctgtcttcccaatgtctcatalgc 1062
Db 648 ACTGAATTTAAAGATGCGGTAGACCACTATTAAGCTGCTTCCCAAGCGCTTTCTGGCT 707
Oy 1063 agtaagctggttcggtgttatagtccctccggtccaggggaggaagctgacctcagtc 1122
Db 708 TCCAAGATGGAACCCGTTGTCTATGTAGAGGATCTCCAGGCTCCAGGCTGACCTGAACCTGC 767
Oy 1123 atggaagactgtctcagagctcagtcggtgtgaaatacttctgaatacattgtggagc 1182
Db 768 ATCAAGATCTTTTGCTCTTCAGAGTCTCATGTGAAGTACGTTATCAACACCTGTGGGCA 827
Oy 1183 gacttccataaagaagaatgcagagatggtccagagctcctcaagatgttgaatggag 1242
Db 828 GACTTCCCCCTGAAACCAACAGAAATAGTTCAAGTATCTGAAGATTTAAAGGTAA 887
Oy 1243 aatgacatgagtcagaggtaccttccctaagcaagaaccgctgtgaatat----- 1296
Db 888 AATATCACCCCAAGGGGTCTGCCCCACCTCATGCAATTTGACGAGACTAAATATGTCCAC 947
Oy 1297 -----cacttgaagtgtagagagacacatataccttaaccacaagaagaagatcct 1350
Db 948 CAAAGCACTCTGGCAAGAGCTTTCTATGTGATGAAGAACACAGCGTTGAACCGCT 1007
Oy 1351 ccccttaataatlaactatgtttacaggaatgctacatgtgtgctcccgagatctc 1410
Db 1008 CCCCCCAATATCTACAAATTTACTTTGGCTGTGCTATGTGCTATCATCAAGAAAGTTT 1067
Oy 1411 gtcaacatgttttgaagaacccttaacccaacatgtatgtgaatggtlaaagaacct 1470
Db 1068 GCCAATCTTGTCTCAATGACCCACAGGGCTGTTGATTTGCTCCAGTGGTCAAGACACT 1127
Oy 1471 tatacccaagatgaacactctggcgccaccttcagcgctgcagcggtgtagtgcctct 1530
Db 1128 TTCAATCTCTGATGACATTTTGGGTGAACACTCAATAGAGATTCCAGGTGTTCTGTGCT 1187
Oy 1531 gtcccaa 1538
Db 1188 ATGCCAA 1195

```

RESULT 10
US-08-474-065-13
Sequence 13, Application US/08474065

```

: Patent No. 5830465
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Minoru
: APPLICANT: Bierhuizen, Marti F.A.
: TITLE OF INVENTION: Expression of the Developmental I
: TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
: TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,065
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/118,906
: FILING DATE: 09-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9526
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1807 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 255..1454
: US-08-474-065-13

```

```

Query Match 7.8%; Score 180.4; DB 2; Length 1807;
Best Local Similarity 54.3%; Pred. No. 8.5e-48;
Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

```

QY	1531	gttcccaa	1538
Db	1188	ATGCCAA	1195
QY	1471	tataagccaagatgaacacacctctggagcaaacaccttcagcgtagcaggtgtagctgacct	1530
Db	1128	TTCACTCCTGATGACACATTCTCTGGGTGACACTCAATAGATATCCAGAGTTCCTCTGCTCT	1187
QY	1411	gtcccaacatgcttttbaaagaaccccttaaaccccaaacacgattgaatgggtaaaagacact	1470
Db	1068	GCCAACTTTGTATCTTCATGACCCACAGGGCGTGTGATTGTGCTCACTGGCAGGAGCACT	1127

```

1 RESULT 11
2 US-08-118-906-3
3 Sequence 3, Application US/08118906
4 Patent No. 5484590
5 GENERAL INFORMATION:
6 APPLICANT: Fukuda, Minoru
7 APPLICANT: Bierhuizen, Marti F.A.
8 TITLE OF INVENTION: Expression of the Developmental I
9 TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
10 TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
11 NUMBER OF SEQUENCES: 14
12 CORRESPONDENCE ADDRESSES:
13 ADDRESSEE: Campbell and Flores
14 STREET: 4370 La Jolla Village Drive, Suite 700
15 CITY: San Diego
16 STATE: California
17 COUNTRY: USA
18 ZIP: 92122
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Floppy disk
21 COMPUTER: IBM PC compatible
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23 SOFTWARE: PatentIn Release #1.0, Version #1.25
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/118,906
26 FILING DATE: 09-SEP-1993
27 CLASSIFICATION: 536
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Campbell, Cathryn A.
30 REGISTRATION NUMBER: 31,815
31 REFERENCE/DOCKET NUMBER: P-LJ 9526
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (619) 535-9001
34 TELEFAX: (619) 535-8949
35 INFORMATION FOR SEQ ID NO: 3:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 378 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: double
40 TOPOLOGY: linear
41 MOLECULE TYPE: cDNA
42 FEATURE:
43 NAME/KEY: CDS
44 LOCATION: 1..378
45 US-08-118-906-3

```

[illegible]

```

RESULT 12
US-08-486-196-3
: Sequence 3, Application US/08486196
: Patent No. 5731420
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Minoru
: APPLICANT: Bierhuizen, Marti F.A.
: TITLE OF INVENTION: Expression of the Developmental I
: TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
: TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/486,196
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/118,906
: FILING DATE: 09-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9526
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 378 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..378
: US-08-486-196-3

```

[illegible]

Query Match	7.0%;	Score 162.8;	DB 1;	Length 378
Best Local Similarity	65.3%;	Pred. No. 1.4e-42;		

[illegible]

OY 1162 tttcgtatcacatgtagagacgccttccataaaagaacaatcgagatgtccaagt 1221
|| ||| |||||| | ||| ||| ||| ||| ||| ||| |||
Db 313 TTGATAAATCTTTGGTGTAGTGATTTCACATTAAACAACCTAGAATATSTCAGAAG 372

OY 1222 ctcaag 1227
||| |||
Db 373 CTCGAAG 378

RESULT 14
US-08-474-065-3
Sequence 3, Application US/08474065
Patent No. 5830465
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,065
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.


```

:      REGISTRATION NUMBER: 31.815
:      REFERENCE/DOCKET NUMBER: P-LJ 9526
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: (619) 535-9001
:      TELEFAX: (619) 535-8949
:      INFORMATION FOR SEQ ID NO: 3:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 378 base pairs
:      TYPE: nucleic acid
:      STRANDEDNESS: double
:      TOPOLOGY: linear
:      MOLECULE TYPE: cDNA
:      FEATURE:
:      NAME/KEY: CDS
:      LOCATION: 1..378
:
US-08-474-065-3

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: September 26, 2002, 03:04:05 : Search time 2719.46 Seconds
(without alignments)
11509.432 Million cell updates/sec

Title: US-09-874-390-1

Perfect score: 2319

Sequence: 1 atactaggggttctcatt.....ataataaacttaacaga 2319

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estb1a:*
2: em_estb2a:*
3: em_estb3a:*
4: em_estb4a:*
5: em_estb5a:*
6: em_estb6a:*
7: em_estb7a:*
8: em_estb8a:*
9: em_estb9a:*
10: em_estb10a:*
11: em_estb11a:*
12: em_estb12a:*
13: em_estb13a:*
14: em_estb14a:*
15: em_estb15a:*
16: em_estb16a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923.6	39.8	1862	11 AK008234	Mus muscu
2	812.2	35.0	864	10 BG821420	BG821420
3	809.8	34.9	826	10 BG821370	BG821370
4	757	32.6	872	10 BG747491	BG747491
5	739.8	31.9	1003	10 BG385575	BG385575
6	711	30.7	875	10 BG470202	BG470202
7	704.8	30.4	766	10 BG386598	BG386598
8	693.4	29.9	866	10 BG468641	BG468641
9	691.6	29.8	731	10 BG248304	BG248304
10	685.6	29.6	702	10 BG386247	BG386247
11	666.2	28.7	920	10 BG248304	BG248304
12	662.2	28.7	920	10 BG386247	BG386247
13	653	28.2	871	10 BG386293	BG386293
14	651.8	28.1	736	10 BE617113	BE617113
15	650.8	28.1	955	10 BG328280	BG328280
16	646.4	27.9	671	10 BG750604	BG750604
17	640	27.6	640	9 AA583339	AA583339

18	633.4	27.3	842	10 BE869192	BE869192
19	622.6	26.8	655	9 AM842622	AM842622
20	615.6	26.5	882	10 BG171085	BG171085
21	613.6	26.5	850	10 BG822181	BG822181
22	613.2	26.4	724	10 BG469448	BG469448
23	611	26.3	612	10 BE292814	BE292814
24	608.4	26.2	912	10 BE871910	BE871910
25	607.2	26.2	1050	10 BE867668	BE867668
26	605	26.1	671	9 AT587061	AT587061
27	599.8	25.9	603	9 AM192901	AM192901
28	597.4	25.8	978	10 BG171515	BG171515
29	586.8	25.3	629	10 BG167918	BG167918
30	582.6	25.1	606	9 AM662377	AM662377
31	557.2	24.0	823	10 BG169857	BG169857
32	556.6	24.0	763	10 B1826675	B1826675
33	552	23.8	659	10 BE304708	BE304708
34	539	23.2	583	9 AM664294	AM664294
35	516.4	22.3	554	9 AA307800	AA307800
36	480.2	20.7	609	10 BE881417	BE881417
37	477.2	20.6	904	10 BG469649	BG469649
38	436	18.8	436	9 AA568218	AA568218
39	433	18.7	433	9 AA583146	AA583146
40	426.4	18.4	433	9 AA315469	AA315469
41	406	17.5	830	10 BE868512	BE868512
42	383	16.5	383	9 A1150400	A1150400
43	381.2	16.4	567	12 A2388491	A2388491
44	377.4	16.3	420	9 AM374998	AM374998
45	375.2	16.2	965	10 BF582709	BF582709

ALIGNMENTS

RESULT 1	AK008234	1862 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK008234				
DEFINITION	Mus musculus adult male small intestine cDNA, RIKEN full-length				
ACCESSION	AK008234				
VERSION	AK008234.1	GI:12842295			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2 (sites)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishikawa, K., Katsunuma, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiki, K., Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				

QY 784 ccttcaagacaccacctaactctccctcaacagagactgtgagcacttcaaggtcaa 843
DB 601 ccttttcacacacacccactactctccctccacagacactgtgacactcaaggctgaa 660
QY 844 aggaagtctacagttccacactgagcaagaaggtgtgagttccctattgatactc 903
DB 661 AGGAAGTCTATACAGTTCCTCCACTGAGCAAGAAGAGGTGAGTTCCTATTGCACTACTCT 720
QY 904 atgtgtatcatagaaagacttgaaaacttgaagagctactgcaagctgtgtatgccct 963
DB 721 ATGCTATTCATGAGAGATGTGAAACTTGAAGGCTACTGCAGCTGTGTATGCTCT 780
QY 964 cagaacatatactgtctcatgtgtgagatgagaagtcctccagaacttcaagaagggcgctc 1023
DB 781 CAGAACATATACCTGTCTCCCTGTGATGAGAAAT-CCCAAGAACTTCCAAACGACGCGG 839
QY 1024 aaagcaatttcttctgtctcccaa 1048
DB 840 CAGCAATATATTCGTGTCCAANA 864

RESULT 3
BG821370 826 bp mRNA linear EST 22-MAY-2001
LOCUS 602724814F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4864634 5',
DEFINITION mRNA sequence.
ACCESSION BG821370
VERSION BG821370.1 GI:14168957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 826)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LICM1728 row: k column: 03
High quality sequence stop: 823.

FEATURES

source
Location/Qualifiers
1..826

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4864634"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 217 a 197 c 205 g 207 t
ORIGIN

Query Match 34.9%; Score 809.8; DB 10; Length 826;
Best Local Similarity 99.6%; Pred. No. 1,6e-224;
Matches 822; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 72 caaggaactgcctgtactctgtgactgcctccttaactcagcagatttctgtcggga 131

DB 2 CACGGGAACAGCCCTGCTACTGTGACCGCCCTTACTACACAGTTTCTTCGGGA 61
QY 132 agcccttgagattctgtcaataactcaactgtagtggctgaaggaagaaacagatgaagac 191
DB 62 AGCCCTGGATTCTGCTAATACCTTATCAGTGTAGTGTCTGAAGGAACAGATGAAGAAC 121
QY 192 atgacctcaagagacttccctgtcaatgagaagaccagctgacgcctggcaaatatata 251
DB 122 ATGACCTCAAGGAGGCTTCCTGTCATGAGAAAGCAAGGTGAGCGCTGGCAAGATATTA 181
QY 252 aagaagagagcctgaaactgttcccttgagacatctatgaatgtcagaataatcctttggag 311
DB 182 AAGAGGAGCCTGGAACCTGTTGACATCTATGAATGCAAAATACCTTTGGAG 241
QY 312 ggttaagaagatcaggagacatgtgtgttcaatttctgcacaggaacagcgaactctt 371
DB 242 GGTTAAGAAGATCAGGGGACATGTTGTTCCATTGCTGCACGGAACCGCAGTCTT 301
QY 372 cacttggaaacagaatcaacgctctgtgaagagatcalccctaagcagagaagaactact 431
DB 302 CACTTGGAAACAGAAATCACGCTTGTGAAGAGATCATCCCTAAGCAGAGAGAACTACT 361
QY 432 aaagagattgtctctccctccactccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 491
DB 362 AAAGGATGTGTCTCTCCCTCCACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 421
QY 492 gaagatgttcaatgagaagagactctgcacagctgcatattgtgtgtgtgtgtgtgtgtgt 551
DB 422 GACGATGCTTCATGAGAGAGAGACTGTGCAGCTGCAATTAATTGTGGGCTGTGGGCTGA 481
QY 552 tatgtctgtgcagactgt 611
DB 482 TATGCTCTGTGCGCACCTGT 541
QY 612 ctgt 670
DB 542 CTGTGGT 601
QY 671 tcttgaacttcagcaagaaggtctcaactcgttcaaggtgtcacccgagggagaccag 730
DB 602 TCTGGAACCTTCCAGCAAGAGGCTATCAACGTTCAAGGGGTACCCGAGGGGACCAAG 661
QY 731 aggcagtgcttcagagctatctgaataaactgtgaaggtcaagaagaagcgagaccttca 790
DB 662 AGGCAGTGTCTCAGGCTTATCTGAATTAACCTGAGGTCAGGAAGAGCGAGACCTTTTCA 721
QY 791 cagaaccacactacactctccctccacagagactgtgaagaccttcaaggtgtgaagaagat 850
DB 722 CAGCAACACACTACCTCTCCCTCAACAGAGACTGTGAGCACTTCMAGCTGAAGAGAGT 781
QY 851 tcatacagttccacactgagcaagaagaaggtgtgagttccctattg 895
DB 782 TCATACAGTTCCTCCACTGAGCAAGAAGAGGTGAGTTCCTATTG 826

RESULT 4
BG747491 872 bp mRNA linear EST 15-MAY-2001
LOCUS 602704606F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4858104 5',
DEFINITION mRNA sequence.
ACCESSION BG747491 GI:14058144
VERSION BG747491.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 872)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L1CM1711 row: k column: 01
 High quality sequence stop: 747.
 Location/Qualifiers

FEATURES

SOURCE

1. 872
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4645159"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 237 a 202 c 214 g 219 t
 ORIGIN

Query Match

Best Local Similarity 32.6%; Score 757; DB 10; Length 872;
 Matches 819; Conservative 0; Mismatches 30; Indels 4; Gaps 4;

185 gaagaacatgacctgaagagcttcctgtcaatgaagaagcgaagcgtgacgacgca 244
 |||||
 Db 1 GAAGAACAATGACCTCAAGAGCTCTCTGCATGAGAACCAAGCTGACGCTGGCANA 60
 245 gattataagagagcgcctgaactgttctcttgacatcttcatgattgtagaataact 304
 |||||
 Db 61 GATTATTAAGAGAGCCTGAACCTCTCTGACATCTTATGAATGTCAGAAATATAC- 119
 305 ttctgaaggttagaagatcaaggagacatgtgttcaactgtctgcacggaacacgc 364
 |||||
 Db 120 TTTCGAGGTTGAGAGATGAGGAGATGTTGTTTGTTCACATTTGCTGCCAGAACACCC 179
 365 cagcttcaacttgaaacagaaatcaacgcctgtgaaagatcatccctaaagcagaga 424
 |||||
 Db 180 CAGTCTTACTTGGAACAGATACAGCCTTGTGAAGAGATCATCCCTAAGCAGAGAGA 239
 425 agctactaaagatgtgtctcctccacacttccctgtgctggtctccactgtctcc 484
 |||||
 Db 240 AACTACTAAAGATGTGCTCCTCCACCTTCCCTGCTGCTGCTCCACCTGCTCTCC 299
 485 attctgtgacgtgttcaatgaagaagacatctgcacgtgattactgtgtggtctcg 544
 |||||
 Db 300 ATTCTGTGACGATGCTTCATGGAAGAGACTCTGCAGCTGATTAATCTGTGGCTTG 359
 545 gctgctataatgtctgacgtgacatgtgctctgaaacttcttcaagttgaagttgact 604
 |||||
 Db 360 GCTGCAATATGCTGCTGGCAGCTGTGGCTCTGAACCTTCTTCAAGTTGAAGTGTGACT 419
 605 ctgacacactgtgtgtgaggtcccaaggaatctcaaacgacgtactgttagaatctgt 664
 |||||
 Db 420 CTGACACACTTGGGTGAGTCCA-GGAATCTCAAGCAGATCTAGGAATATCTTGT 478
 665 ataattctgaaacttcaagcaagaggttatcaactgttcaagggtcaccgagggg 724
 |||||
 Db 479 ATATATTTCTGAAACTTCACAGCAAGAGGTCTATCACTGTTCAAGGGGTCAACCGAG 538
 725 accaagaagcagtgcttcaagctatcttgaataaactggaaggtcaagaagaagcgaagc 784
 |||||
 Db 539 ACCAAGAGGACGTGCTTCAAGCTATCTGAATACCTGAGAGTCAAGAAGAGCGAGAC 598

Qy 785 ctccacagacaccactactctccctcaccagagactgttagacttcaaggtcga 844
 |||||
 Db 599 CTTTCACAGACACCCACAT-CTCTCCCTACACAGACAGCTGATGACTTCAAGGCTGAA 657
 Qy 845 ggaagttatacagttccactgaagcaagaagaggtgagttccattactactcta 904
 |||||
 Db 658 GGAAGTATATACAGTTCCACATGACCAAGAGAGGTGAGTCCCTATTCATCTACTGA 717
 Qy 905 tggatctatcagaagaagattgaaacttgaagagctactgtgagctgtgtat-gccct 963
 |||||
 Db 718 TGGTATTCATGAGAGATGAAGCTTGAAGGCTCTGAGAGCTGTGATGGCCCT 777
 Qy 964 cagaacataatgtgtcatgtgtagaagtcaccagaacttcaagagcggtc 1023
 |||||
 Db 778 CAGAAATATTTACTGGGTCTCTGTGATTGAGAGTTCCCCCAAAATTTCAAAAAGGGGGC 837
 Qy 1024 aaagcaattattt 1036
 || || ||
 Db 838 AAGCAGATATTTT 850

RESULT 5
 BG469141
 LOCUS 602511125P1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4645159 5',
 DEFINITION mRNA sequence.
 ACCESSION BG469141
 VERSION BG469141.1 GI:13401416
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 912)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L1CM1418 row: b column: 08
 High quality sequence stop: 705.
 Location/Qualifiers

FEATURES

SOURCE

1. 912
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4645159"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 240 a 214 c 231 g 227 t
 ORIGIN

Query Match 32.6%; Score 757; DB 10; Length 912;
 Best Local Similarity 94.2%; Pred. No. 4.2e-209;
 Matches 841; Conservative 0; Mismatches 45; Indels 7; Gaps 5;

```

QY 74 cggagactgacctgtactgtgacactgaccttaactcagcagcttctgtctggag 133
Db 2 CGGGAACCTGCCCTTGCTACTGTTGACCTGCCCTTACTCAGCACTTTTGTTCGGGAAG 61
QY 134 cccctgacttctgactaatactactgactgtgactgactgaggaagaagaagaacat 193
Db 62 CCCGGAATTCCTGCTAATACCTATACCTGTAAGTGTGAAGGAACAGATGAAGAACAT 121
QY 194 gacctcaaggagctctctgtcaatgagaagcaagaactgagccttggcaaatataa 253
Db 122 GACCTCAAGAGAGCTTCTGCTCAATGAGAACCAAGCTGAGCCTGGCAAAAGATATATAA 181
QY 254 gaggagcctgaaactgttctctgtgacatctatgaaatgacaaataactcttggagg 313
Db 182 GAGAGAGCTGAACCTGTTCTCTGAGACATCTTATCAATGTGAGAAATACCTTTTGGAGGG 241
QY 314 ttagaagatacagagagagactgtgttcaacattgtctgcaagagacccgacattca 373
Db 242 TTAAAGATCAGGGGACATGCTTTTCAATTTCTGCGACGGAACACCGGACATCTTCA 301
QY 374 ctggaaacagaatcaacgacctgtgaaagagatacaccctaagcagagagaagactataa 433
Db 302 CTTGAAACAGAAATCAGCGCTTGTGAAGAGATCACCCTAAGCAGAGAGAACTACTAA 361
QY 434 aggatgtgtctctctctcaactctccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 493
Db 362 AGGATTTGTCTCTCTCTCACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
QY 494 cgaagtgtcaatgagaagagactctgcaagcagcagcagcagcagcagcagcagcagcag 553
Db 422 CGATGCTTCAATGAGAGAGACTCTGCCAGCTGCAATTTGTTGGGCTTGGGCTGCTATA 481
QY 554 tgcgtgtgacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 613
Db 482 TGCTGCTGCGCAGCTGTGCTGTGAACCTTCTTCAAGTTGAGTTGAGTGTGACTGACACAT 541
QY 614 tgggtgtgagatcaggaagatctcaaaagcagctactgtgagaatatactgtataatttc 673
Db 542 TGGGTGTGAGATCCA-GGAATCTCAAGCCAGTACTGAGGAATATCTTGTATATATTTCC 600
QY 674 tgaactctcagcaagaagagctcactcaactgttcaaggggtacaccgagaggaagag 733
Db 601 TGAACCTTCCAGCAAGAGGCTATCAACTTTCAGGGGTACCCGAGGGAGCAAGAGG 660
QY 734 cagtgtctcaggtctactctgaaataactgtgaaaggtcgaagaagcgaagccttcaag 793
Db 661 CAGTGCTTCAAGCTATCTGAAATACCTGGAGGTCAAGAAAGCAGAGAGCTTTTTCACAG 720
QY 794 aaccccaactcctctccctcaacagaga--ctgtgagcacttcaaggtcgaagaagat 851
Db 721 ACACCAATAGCTTACCTCTACACAGAGAACTGTGAGGCACTTCAAGGCTGAAAGGAGCT 780
QY 852 c-atacagttccactgagcaagaagaggtgtgagttccctatgctactactatgtg- 909
Db 781 CAATACAGTTCCACTGAGC-AAGAAGAGGTGAGGTTCCATATAGCATACTCTATGGGGG 839
QY 910 -atcatgagagaattgaaacttgaagaactgtcgaagctgtgtatgtcc 961
Db 840 AACCATGAGAGCATGTGAACCTTGGAAGGCTATGGAAGCTGTGAAGACC 892

RESULT 6
LOCUS BG385575 1003 bp mRNA linear EST 12-MAR-2001
DEFINITION 602453872F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582504 5',
            mRNA sequence.
ACCESSION BG385575
VERSION BG385575.1 GI:13278389
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE 1 (bases 1 to 1003)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
        Email: cgabs-r@mail.nih.gov
        Tissue Procurement: ATCC
        cDNA Library Preparation: Ling Hong/Rubin Laboratory
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
        DNA Sequencing by: NIH Intramural Sequencing Center
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNL at:
        http://image.lnl.gov
        Plate: LNCMI304 row: 0 column: 17
        High quality sequence stop: 694.
        Location/Qualifiers
            1..1003
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4582504"
                /clone_lib="NIH_MGC_15"
                /tissue_type="adenocarcinoma cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: colon; Vector: pORF7; site_1: XhoI; site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCAAGAG(s). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 282 a 248 c 264 g 209 t
ORIGIN
Query Match 31.9%; Score 739.8; DB 10; Length 1003;
Best Local Similarity 93.1%; Pred. No 4.6e-204;
Matches 818; Conservative 0; Mismatches 57; Indels 4; Gaps 4;
QY 174 gggaaacagatgaagaacatgacctcaagagactctctgtcaatgagaagcaagctga 233
Db 1 GGGAAACAGATGAGAACATGACCTCAAGAGGCTTCTGCAATGAGAACCAAGCTGA 60
QY 234 cgcctggcaagaatataaagagagcctgaagaactgttctctgtgacatctatgaaatgc 293
Db 61 CGCTGGCAAAAGATTAATGAAGAGAGCGCTGAACCTGCTTGGACATCTTATGAATGTC 120
QY 294 aagaataaccttctgagaggttagaagatcagaggagacatgtgttcaatttgcgca 353
Db 121 AGAAATTAACCTTTGGAGGTTAGAAATCAAGGGAATGTTTCAATTGCTGCCA 180
QY 354 cggacacccgcagctcttcaacttgaacacagaatcaacgcttgtgaagatcatcccta 413
Db 181 CGGAACACCGCAATCTTCACTTGAAGAAAGAAATACAGCGCTTGTGAAGAGATCATCCCTA 240
QY 414 agcagagagaagctactaaagagattgttctctccctccacttccctgtgtgtgtcc 473
Db 241 AGCAGGAGAGAAGCTACTAAAGATGTGTCTCTCCACTTCCCTGTCGTGCTGCTGCC 300
QY 474 acctgtctccattctgtgacagatgttcaatgagaagactctgcaagctgatactt 533
Db 301 ACCTGTCTCCATTTCTGTGACGATGTTCAATGGAAGAAAGACTGTGCCAGCTGCAATTA 360
QY 534 gtggagctctggagctcatatgtctgtcggcgaactgtgtgctgaacttcttcaagtc 593
Db 361 GTGGGCTCTGGGCTGCTATATGCTGCTGCGCACCTGTGCTGTAACCTTCTTCAAGGT 420
QY 594 gaagtgtactctgacacttgggtctgtgagctcagaggaattcgaagccagactgtag 653
Db 421 GAACTGTGACTCTGACCACTTGGGTGAGTCCAGGGAATTCAAAGCAATGACTAG 480
QY 654 gaatacttgtataattctctgaaactccaagcaagaggtctataactgttcaagggt 713

```

Db 481 GAATATCTTGATATATTTCTGAAACTTCAGACAAAGAGTCTATCAACTGTTCAGGGGT 540
Qy 714 caccgagagggaccagagagcagtcctcagcgtatctcgaataaaccctggaggtcaaga 773
Db 541 CACCCGAGGGGACCAAGAGGCAAGTCTTCAGGCTATCTTGATTAACCTGGAGGTCAAGAA 600
Qy 774 gaagcgaagacgtctcacaagacacccactacgtctccctccacagagactgtgaacatt 833
Db 601 GAAGGAGAGGCTTTCACAGACACCCACTTACTCTCCCTCCAGACAGACTGGAGACTT 660
Qy 834 caagcgtcaaggaattcacaagctccacactgagcaagaagaggtgtgagttcccat 893
Db 661 CAAGGCTCAAGAGAGTCTCATACAGTT-CCACTGAGCACAGCAGGTGAGTTCCTCAT 719
Qy 894 tgcct-actctatgtgtatcatagaa-gattgaacatttgaaggtactgtgaact 951
Db 720 TGCATAACTCTATGGGATTCACGAAAGCTTGACACTTTCAGAGCTTAATGGAGCT 779
Qy 952 gtgtatgcccctcag-aacataactgtgtccatgtgagatgagaaagtcgccagaacatt 1010
Db 780 GGGATGTACCTCAAAACATATACGGGGCCCGGTGGCATGAAGAGCCCGAAATTC 839
Qy 1011 caaagagcggtcaagcaattattctgtctcccaa 1049
Db 840 ACAGAGGGGGCAACACATATCTGAGGTCCCAA 878

RESULT 7
Bg470202 875 bp mRNA linear EST 21-MAR-2001
LOCUS 602533711F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4661355 5',
DEFINITION mRNA sequence.
ACCESSION Bg470202
VERSION Bg470202.1 GI:13402477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 875)
NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rsb@ncl.nih.gov
Tissue Procurement: AMCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LUCM1460 row: e column: 04
High quality sequence stop: 814.
Location/Qualifiers
1..875
/organism="Homo sapiens"
/db_xref="taxon.9606"
/clone="IMAGE:4661355"
/clone_id="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 223 a 211 c 219 g 222 t
ORIGIN

Query Match 30.7%; Score 711; DB 10; Length 875;
Best Local Similarity 93.5%; Pred. No. 1,1e-195;
Matches 811; Conservative 0; Mismatches 40; Indels 16; Gaps 6;
Qy 391 gctctgtgaagatcattccctaaagcagagagaaagacttaaaagattgttcctc 450
Db 2 GCTTGTGAAGATCATCCCTAAGCAGAGAGAAAGTACTTAAGATGTGTCTCTCC 61
Qy 451 caacttccctgtgtcgtctccaccctgtcccaattctgtgaagaatgttcaatgaag 510
Db 62 CACTTCCCTGTGCTGGCTTCACACTGTCTCCATTCGTGACGATGGTCAAGGAAG 121
Qy 511 agactctgcagctgacttactgtgtggtctgtggtctctatgtctgtggtggt 570
Db 122 AGACTTGCACAGCTCATTAAGTGTGGCTGTGGCTGTCTATAGTCTCTGGCACTGTG 181
Qy 571 gctctgaacttctcttcaagttgaaagtgtactctgacactgtgtctgtgagtcag 630
Db 182 GCTTGAAACTTTCTTTAGGTGAAGTGTGACTGTGACCACTTGCTGGAGTCCAGG 241
Qy 631 gaattcagaagcagctactgtgaagaaatctgtataatttccgtgaactccagaag 690
Db 242 GAATCTCAAGCCAGTACTGTAGAATATCTGTATATTTCTGAAACTTCCAGCAAG 301
Qy 691 aggtctcaactgttcaaggggtccacccgagagagcagaagaggtgtctcaggct 750
Db 302 AGGTATATCAACTGTTCAGGGGTACCCGAGGGGACCAAGAGGAGTCTTCAGGCTAT 361
Qy 751 ctgaataaccctggaagtgcaagaagaagcagagccttccacagaaacacactactc 810
Db 362 CTGAATTAACCTGGAGGTCAAGAGAGAGAGAGCCTTTCACAGACACCACTACTCT 421
Qy 811 ctacacagagactgtgagcaacttcaaggtctgaagaaatctacagattcccaatg 870
Db 422 CTCACAGAGAGCTGTGACACTTCAAGGCTGAAAGAGTTCTATCACTTCCACTGAG 481
Qy 871 aaagaaggggtgaggttccctatctgcatctctatgtgattcatgagaagattga 930
Db 482 AAGAGAGGCTGAGTTCCTTATTCATATCTATGATGATTCATGAGAGATTGAAC 541
Qy 931 ttgaaagctactctgagactgtgtatgtccctccagaaacataactgtgtccatg 990
Db 542 TTGAAAGGTACTCTGAGGTGTGTATGCTCCCTCAGAACATATGATGTGATGTGAT 601
Qy 991 gagaagtcctccagaacatttcaagaagcgttcaagaagaattatttctgtctcc 1050
Db 602 GAGAAGTCCCAAGAACTTCAAGAGGCGGTCAAGCAAGCAATTTCTTCCCAAT 661
Qy 1051 gttctcatagccagtaagctgtgtgtgtttat---gctcctgtgtccaggtgca 1107
Db 662 GTCTTCATAGCCAGTAACTGCTGTGGGTGTATACCTTCCTCCAGGCTGCA 721
Qy 1108 ---gttgcctcaactgtatgaa--gaattgtccagaagctcagtgccgtggaata 1161
Db 722 AGCTTGACCTTCACATGCTGAGAAAGACTTGGCTCAGAGCTCACTGCGTGAATAC 781
Qy 1162 ttctcgaatacaatgtggg-----acggaacttccataaagaagcaatgag--aga 1214
Db 782 TTCTGGAATACCTTGTGGAAACCGGACTTTCTTATTAAGAGCAATGACAGATG 841
Qy 1215 caagcgtcagaagtgtgaatggag 1241
Db 842 CCAGGCTCAAGATGTGACACGGGG 868

RESULT 8
B1765410 766 bp mRNA linear EST 25-SEP-2001
LOCUS B1765410
DEFINITION 603050445F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5190611 5',
ACCESSION B1765410
VERSION B1765410.1 GI:15756988

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 766)
AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Hominae: Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Inceye Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11476 row: 1 column: 12
High quality sequence stop: 766.

FEATURES
source
1. 766
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5190611"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector: PCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."

BASE COUNT 199 a 184 c 192 g 190 t 1 others

ORIGIN

Query Match 30.4%; Score 704.8; DR 10; Length 766;
Best Local Similarity 98.4%; Pred. No. 6.3e-194;
Matches 753; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 316 agaaagatcaggagacatggtgttcacattgtgtccaggaacacccgacatcttact 375
DB 1 AGAAGATCAGGNGACATG--TGTTACATTTGCTGCCAGGACCCGACGCTTCAGG 59
QY 376 tggaaacagaatcagcctgtgaagatcatccctaaagcagagagaactactaaag 435
DB 60 GGGAAACAGATCAGCGCTGTGAAGAGATCCCTAAAGAGAGAGACTACTAAAG 119
QY 436 gattgtctctctcaccttcctgtgtcgtgttcacattgttccatttctgtgagc 455
DB 120 GAGTGTCTCTCTCCACCTTCCTGTGCTGCTCCACCTCTCTCCATTCGTGACG 179
QY 496 atggttcaatggaagacatctgcacatctactgtgtggctctgggctgtatag 555
DB 180 ATGTTCAATGGAAGACATCTGCCAGCTCATTTGCTGCGCTCGGGCTGCTATATG 239
QY 556 ctgctggccactgtgtgctgaacttcttcaagtlgaagtgtgactgtgacactg 615
DB 240 CTGCTGGCCACTGTGGCTGGAACCTTTCTTTCAGGTGAAGTGTGACCTGACCACTTG 299
QY 616 gttcttgaggtccaggaatctcaagccagctactgttagaatactctgtataattctg 675
DB 300 GGTCTGGAGTCCAGGAGATCTCAAGCCACTACTGTGAATATCTTGTAAATTCCTG 359
QY 676 aaattcccaagaagaagttatcaactgttcagggtgtcccgaggggcgaagagga 735
DB 360 AAACCTTCACGAAAGAGTCTATCACTGTTCAGGGGTACCCGAGGGGACCAAGAGGCA 419

QY 736 gtgttcaggctatcttgaataaccctggaaggtcaagaagaagcgaagccttcaagac 795
DB 420 GTGCTTCAGGCTATTCGTGAATACCTGAGAGTCAAGAGAGGAGCCTTCACAGAC 479
QY 796 acccaactactcttcctccaccgaagactgtgaacacttcaagcgtgaaggaattcata 855
DB 480 ACCCACTACTCTCTCCCTCACAGAGACTGTGACACTTCAAGGCTAAAGAGATTCTATA 539
QY 856 cagttcccaatggaagaagaagaggtggaattccctattgacactctatgtgtatcat 915
DB 540 CATTTCCTCATGAGCAAGAAGAGGTGAGACTTCCATTGCTCATGCTATGAGATTCAAT 599
QY 916 gagaagaattgaaaacttgaagaagctactgagactgtgtatgtccctca-gaacaata 974
DB 600 GAGAAGATTGAAAACCTTTGAAAAGCTACTGCGAGCTGTATGCCCCCTCAGAAACATATA 659
QY 975 ctgt-gtccatgtgagtgagaatcccccgaagacttcaag-aggcgttcaagaagaatt 1032
DB 660 CTGTGTCCATGTGATGAGAGAGTCCCGACACTTTCAAAGAGCGGGTCAAGACATAAT 719
QY 1033 atttctgtctcccaaatgtcttcatagccaagtaagctgttcg 1077
DB 720 ATTTCTTGCTTCCCAATGATCTTCATAGCCAGTAACTGTGCGG 764

RESULT 9
BG386598 866 bp mRNA linear EST 12-MAR-2001
LOCUS 602454847F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583243 5',
DEFINITION mRNA sequence.
ACCESSION BG386598
VERSION BG386598.1 GI:13280044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM1306 row: n column: 12
High quality sequence stop: 744.

FEATURES
source
1. 866
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583243"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 219 a 209 c 221 g 217 t

Query Match 29.9%; Score 693.4; DB 10; Length 866;
 Best Local Similarity 97.9%; Pred. No. 1.4e-190;
 Matches 755; Conservative 0; Mismatches 11; Indels 5; Gaps 5;

442 gtccctccaccctccctctgtgtgtcctcaactgttcccatctgtgacgatgtt 501
 |||||
 Db 2 gtccctccaccctccctctgtgtgtcctcaactgttcccatctgtgacgatgtt 501
 |||||
 502 caatgaagaagctgtccagctgcatctactgtgtgtgtgtgtgtgtgtgtgt 561
 |||||
 Db 62 caatgaagaagctgtccagctgcatctactgtgtgtgtgtgtgtgtgtgtgt 121
 |||||
 562 gccactgt 621
 |||||
 Db 122 gccactgt 181
 |||||
 622 ggtgtcaggggaatctcaaaagcagctgtgtgtgtgtgtgtgtgtgtgtgt 681
 |||||
 Db 182 ggtgtcaggggaatctcaaaagcagctgtgtgtgtgtgtgtgtgtgtgtgt 240
 |||||
 682 caagcaagaaggt 741
 |||||
 Db 241 caagcaagaaggt 300
 |||||
 742 caggtatctgtgaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 801
 |||||
 Db 301 caggtatctgtgaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360
 |||||
 802 taactctccctcaagaagactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 861
 |||||
 Db 361 taactctccctcaagaagactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
 |||||
 862 caactgaagaagaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 921
 |||||
 Db 421 caactgaagaagaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
 |||||
 922 attgaagaactgtgaagaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 981
 |||||
 Db 481 attgaagaactgtgaagaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
 |||||
 982 catgt 1040
 |||||
 Db 541 catgt 600
 |||||
 1041 ctccccaagtgtctcatagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1099
 |||||
 Db 601 ctccccaagtgtctcatagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 660
 |||||
 1100 ggtgtgaagctgtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1158
 |||||
 Db 661 ggtgtgaagctgtgtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
 |||||
 1159 taactctcgaatacatgtt-gggagcagctgtgtgtgtgtgtgtgtgtgtgt 1208
 |||||
 Db 721 taactctcgaatacatgtt-gggagcagctgtgtgtgtgtgtgtgtgtgtgt 771
 |||||

RESULT 10
 BG468641

LOCUS BG468641 731 bp mRNA linear EST 21-MAR-2001
 DEFINITION 605310243p1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4644467 5',
 mRNA sequence.
 ACCESSION BG468641
 VERSION BG468641.1 GI:13400911
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 731)
 AUTHORS NIH-MGC http://mgi.mgi.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.lnl.gov
 Plate: LUCM1416 row: e column: 12
 High quality sequence stop: 686.
 Location/Qualifiers
 1..731
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:464467"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOMB7; Site: 1: XhoI; Site: 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using 2AP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 214 a 157 c 183 g 177 t
 ORIGIN

Query Match 29.8%; Score 691.6; DB 10; Length 731;
 Best Local Similarity 98.4%; Pred. No. 4.3e-190;
 Matches 720; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

568 gtgtgtctgtgaacttctcttcaagtgtaagtgtaactgtgacactgtgtgtgtgt 627
 |||||
 Db 2 gtgtgtctgtgaacttctcttcaagtgtaagtgtaactgtgacactgtgtgtgtgt 61
 |||||
 628 aaggaatctcaagaagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 687
 |||||
 Db 62 aaggaatctcaagaagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 121
 |||||
 688 aaggaatctcaagaagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 747
 |||||
 Db 122 aaggaatctcaagaagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 181
 |||||
 748 attctgaataaactgt 807
 |||||
 Db 182 attctgaataaactgt 241
 |||||
 808 tccctcccaagaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 867
 |||||
 Db 242 tccctcccaagaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 301
 |||||
 868 agcaagaagaaggt 927
 |||||
 Db 302 agcaagaagaaggt 361
 |||||
 928 aacttgaagaagctgt 987
 |||||
 Db 362 aacttgaagaagctgt 421
 |||||
 988 gatgtgaaggt 1047
 |||||
 Db 422 gatgtgaaggt 481
 |||||
 1048 aatgtcttcataagcagtaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1107
 |||||
 Db 482 aatgtcttcataagcagtaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 541
 |||||
 1108 gctgaactcaactgtgtgaagaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1167
 |||||

Db 542 GCTGACCTCACTGATGGAAGACTTGGTCACAGCTACGTGCCGTGGAATACTTCTG 601
 Qy 1168 aatacatgttggaacgagcttccataaagaacgaatgcagagatggtccagctcctaag 1227
 Db 602 AATACATGCTGGAGGAGCTTTCCTATTAAGACGATGCAAGATGCTCCAGG--CTCAG 659
 Qy 1228 atgttgatgggaagaaatgacatgagat--cagagtcctcctcctaagcacaagaacccg 1286
 Db 660 ATGTTGAAATGGAGGAATGACATGAGTACAGAGGTAACTCATTAAGCAAAAAGAAAAACG 719
 Qy 1287 ctggaatatca 1298
 Db 720 TTGGAATATATA 731

RESULT 11

BG248304 702 bp mRNA linear EST 13-FEB-2001
 LOCUS 602400340F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546017 5',
 DEFINITION mRNA sequence.

ACCESSION BG248304
 VERSION BG248304.1 GI:12758119
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (Bases 1 to 702)
 NIH-MGC http://mgc.nci.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM1231 row: o column: 10
 High quality sequence stop: 702.
 Location/Qualifiers

FEATURES

source

1..702
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4546017"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Origin: Vector: POTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 186 a 163 c 175 g 178 t
 ORIGIN

Query Match 29.6%; Score 685.6; DB 10; Length 702;
 Best Local Similarity 99.4%; Pred. No. 2,4e-188;
 Matches 688; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 119 ttgttgctgggaagccttgatctgctaatacctatactgtaagtctgaaggaa 178
 Db 1 ttttggtctgggaagccttgatctgctaatacctatacctgtaagtctgaaggaa 60
 Qy 179 acagataagaacatgacccatgaagagcttccgttcaatgagaagaccagctacgcct 238
 Db 61 ACAGATGAAGACATGACCTCAAGAGAGCTTCCCTCTCATATGAGAAGACCAAGCTACGCCCT 120

Qy 239 ggcacaagatacctaagaagagccttgaaactgttcccttgagacatcattgaatcagaa 298
 Db 121 GGCAAAGATATTTAAAGAGGAGCGCTGAAGACTTCTTGGACATCTATGATGATGAGAA 180
 Qy 299 ataccctttggaaggttaagaagatcaaggagacatggtgtgttcacattgtgtgcacgaa 358
 Db 181 ATACCTTTGGAGGGTTAGAGATCAGGGACATGGTTGTTCAATTTGCTGCCACGGAA 240
 Qy 359 caccgcagctcctcaacttgaaaacagatacagcctctgttgaagaagatcaccctaaagcag 418
 Db 241 CACCGCCAGTCTTCACTTGGAACAGATCAGCGCTTGTGAAGAGATCATTCCCTAAGCAG 300
 Qy 419 gagaagaagactaaagaatgtgtctcctcctcaactcctctgtgtcgtgtccacatg 478
 Db 301 GAGAGAAGCTACTAAAGAGATTGTCTCTCCACCTTCCCTGGCTGCGGTCCACTG 360
 Qy 479 ttcccatcttgtagagatgttcaatgagaagacctgcagcgtgcattactgtggg 538
 Db 361 TCTCCCACTTCTGTGACATGTTCAATGAAGAGACTCTGCAGCTGCATTACTGTGGG 420
 Qy 539 cctcgtgctgctatagtctgtgcccactgtgctcctgaactctcttcaagttgaagt 598
 Db 421 CTCTGGGCTCTATATGCTGCTGCGCACTGTGCTGTGAACCTTTCTTACAGTTGAGT 480
 Qy 599 gtgacttcgacacttggtgtcgtgagtcaggaatcctaaagccagctactgtagaata 658
 Db 481 GTGACTCTGACCACTTGGGTGCTGAGTCCAGGGAATCTCAACCCAGTACTGTAGGATA 540
 Qy 659 tcttgataattcttcgaactccagcaagaggttctcaactgttcagggttacc 718
 Db 541 TCTTGATTAATTTCTTAACCTTCACGAAAGAGGTATCAACTGTTCAAGGGGTACCC 600
 Qy 719 gaagggaacaaagagcagtgcttcaaggtcattctgaataaccctggaagaaagagc 778
 Db 601 GAGGGAGCAAGAGGCGCTCTCAGGCTATGTCGATTAACCTGAGAGTCAAGAAAGC 660
 Qy 779 gagaagccttcaagaagacccactactctcc 810
 Db 661 GAGAGCCTTTCACAGAAACCATTAACCTCTCC 692

RESULT 12

BG386247 920 bp mRNA linear EST 12-MAR-2001
 LOCUS 602454539F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583984 5',
 DEFINITION mRNA sequence.

ACCESSION BG386247
 VERSION BG386247.1 GI:13279693
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (Bases 1 to 920)
 NIH-MGC http://mgc.nci.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM308 row: m column: 09
 High quality sequence stop: 749.
 Location/Qualifiers

FEATURES

source

1..920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4583984"

/clone_1lb="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 249 a 210 c 238 g 223 t
ORIGIN

Query Match 28.7%; Score 666.2; DB 10; Length 920;
Best Local Similarity 93.3%; Pred. No. 1.2e-182;
Matches 751; Conservative 0; Mismatches 48; Indels 6; Gaps 5;

QY 174 ggaagacagatgaagaacatgaccccaaggagcttcctgtaacatgagaagaccagctga 233
DB 2 gggaaacagatgaagaacatgaccccaaggagcttcctgtaacatgagaagaccagctga 61
QY 234 cgctgagcaagaatataaagaagagagctgaactgttccttgagacatcctatgaatgc 293
DB 62 cgccctggcaagatatttaagagagagctgaactgttccttgagacatcctatgaatgc 121
QY 294 agaaatacctttggagaggttaagaagaacaggagacatgtttcaccatttgcctgca 353
DB 122 agaaatacctttggagaggttaagaagaacaggagacatgtttcaccatttgcctgca 180
QY 354 cggaaacacgcagctcctcacttggaacacgaatacagccttgttgaagaatcacccta 413
DB 181 cggaaacacgcagctcctcacttggaacacgaatacagccttgttgaagaatcacccta 240
QY 414 agcagagagagaagctactcaaaagatgtgtccctccacacttccctgtgctgcctcc 473
DB 241 agcagagagagaagctactcaaaagatgtgtccctccacacttccctgtgctgcctcc 300
QY 474 accctgccacacttgtagagatgttcaatgaagaagagacttgcagctgagcttactt 533
DB 301 accctgccacacttgtagagatgttcaatgaagaagagacttgcagctgagcttactt 360
QY 534 gtaggctctgggctgctatgctgctgctgctgctgctgctgctgctgctgctgctgct 593
DB 361 gtaggctctgggctgctatgctgctgctgctgctgctgctgctgctgctgctgctgct 420
QY 594 gaagtgtgactctgacacacttggtgtgagtgatcgaaggaaatcgaagccagctagtga 653
DB 421 gaagtgtgactctgacacacttggtgtgagtgatcgaaggaaatcgaagccagctagtga 480
QY 654 gaatacttgtataaatttcctgaacacttcgaagaagagagcttcaactgttcagggt 713
DB 481 gaatacttgtataaatttcctgaacacttcgaagaagagagcttcaactgttcagggt 540
QY 714 c-acccgagggagacaaagagcagtgcttcaagagctatctgaataacttgagagctcaaga 772
DB 541 caaccggagggagacaaagagcagtgcttcaagagctatctgaataacttgagagctcaaga 600
QY 773 agaagcgaagagccttcaagacacacacttccctcaacacagagactgtgagcaat 832
DB 601 agaagcgaagagccttcaagacacacacttccctcaacacagagactgtgagcaat 660
QY 833 t-caagtgtaagaagaattcatatagttccacacgagcaagaagaagtgtagtccct 891
DB 661 tccaaagggtgaagaagaattcatatagttccacacgagcaagaagaagtgtagtccct 720
QY 892 attgcatcct-ctatggtgattcatgagaga--ctgaagaacttggaaggtcattgca 948
DB 721 attgcatcctcctatggtgattcatgagagaagaattgcaaacacttggaaggttctc 780
QY 949 gctgtgtagtccctcagaacatat 973

DB 781 ACTGGGTTGGCCTGAGAACATATCT 805

RESULT 13
LOCUS BG386293 871 bp mRNA linear EST 12-MAR-2001
DEFINITION 602455486F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583948 5',
mRNA sequence.
ACCESSION BG386293
VERSION BG386293.1 GI:13279739
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LNCM1308 row: k column: 21
High quality sequence stop: 736.
Location/Qualifiers

FEATURES
source 1..871

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583948"
/clone_1lb="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 231 a 212 c 215 g 213 t
ORIGIN

Query Match 28.2%; Score 653; DB 10; Length 871;
Best Local Similarity 95.2%; Pred. No. 8.5e-179;
Matches 793; Conservative 0; Mismatches 25; Indels 15; Gaps 11;

QY 64 ctctaagtcacggaggaactgacctgtactgtgacagctgaccttactcagcagttttg 123
DB 2 ctctaagtcacggaggaactgacctgtactgtgacagctgaccttactcagcagttttg 60
QY 124 ttcttggaagcccttggaattctgctaactactacatgtagtgctgaaaggaaacga 183
DB 61 ttcttggaagcccttggaattctgctaactactacatgtagtgctgaaaggaaacga 120
QY 184 tgaagaacatgaacctcaaggagcttcctgtaataagaagaacagctgaagccttgga 243
DB 121 tgaagaacatgaacctcaaggagcttcctgtaataagaagaacagctgaagccttgga 180
QY 244 agatatlaaagagagccttgaaactgttccttgagacatcttatagaatgtcagaataatcc 303
DB 181 agattttaagagagagccttgaaactgttccttgagacatcttatagaatgtcagaataatcc 240
QY 304 ttcttgagaggttagaagatcagaggagacatggtgtgtcaccattgtcgcacgaagacgcg 363
DB 241 ttcttgagaggttagaagatcagaggagacatggtgtgtcaccattgtcgcacgaagacgcg 300

QY 364 ccagcttcctccttggaacagaatcaacgcgcttgtaagagatcatccctaagcagagag 423
 |||||||
 Db 301 CCAAGTCTTCACTTGGAACAGATCAAGCCTTGTAAGAGATCATCCCTTAAGCAGAGAG 360
 QY 424 aagctactaaagatgtgtcctcctccaccttcctcctgtgtcgtctccacctgtctcc 483
 |||||||
 Db 361 AACCTACTAAAGATGTGTCTCTCCACCTTCCCTGTCTGTCTCCACCTGTCTCC 420
 QY 484 cactctgtgacgagtgttcacat--ggaagagactctgcccgcgccttacttctgtgctc 541
 |||||||
 Db 421 CATTCTGTGAGATGGTTCATCGGGAAGAGACTGTGCCAGTGCATTTACTTGTGGGCTC 480
 QY 542 tgggtctgctatactgtcctgtgcacatgtgtccttgaaacctcttccaagttgaagtggt 601
 |||||||
 Db 481 TGGGCTGCTATATGCTGCTGCGCAGCTGTGCTGGAACCTTTCTTTCAGGTTGAAGTGTC 540
 QY 602 actctgacacactgtgtgtcgtgagtcacagggaaatcacaagccagctactgtagaatact 661
 |||||||
 Db 541 ACTCTGACCACTTGGGTCTGGAGTCCA--GGAATCTCAAAAGCCAGTACTGTAGGAATATCT 599
 QY 662 tgaataa-cttctctgaacttcacagc-aagaagagttcataa-ctgtctagggtgtcacc 718
 |||||||
 Db 600 TGTATATTTTCTTGAACACTTCCAGCAAAAGAGGTCTATCAACCTGTTCAGGGGTCAACC 659
 QY 719 ---gaggggacacagagagcagtgctt--caggtctatctgaataacctgtgaagttca-aga 772
 |||||
 Db 660 CGAGGGGACCAAGAGGCGAGTGTTCAGGGTATCTGATTAACCTGGAGGTCACAGA 719
 QY 773 agaaagagagagcctttcac-agaacacacactactcctcctcaccagagaagctgtgagc-a 830
 |||||||
 Db 720 AGAAGCGAGGCGCTTTCACAAAGACACCCACTACTCTCTCAACAGAGCTGGAGCAAA 779
 QY 831 ctccaaggtcgaaagagagcttaccagcttcacagctgaagcaagaagaaggttg 883
 |||||||
 Db 780 CTTCAGAGCTGAAGGAGACGTATACCACTCCCATAGCAGCAAAAGAAAGTGG 832

RESULT 14
LOCUS BE617113 736 bp mRNA linear EST 20-OCT-2000
DEFINITION 601441649F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845677 5',
 mRNA sequence.
ACCESSION BE617113
VERSION BE617113.1 GI:9888051
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabds-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: L1AM9557 row: b column: 14
 High quality sequence start: 9
 High quality sequence stop: 670.
FEATURES
 source
 1..736
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3845677"
 /clone_1ib="NIH_MGC_65"
 /tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCW-SPOrt6; Site: 1: NotI;
 Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

BASE COUNT 191 a 170 c 183 g 191 t 1 others
ORIGIN

Query Match 28.1%; Score 651.8; DB 10; Length 736;
Best Local Similarity 99.3%; Pred. No. 1.7e-178;
Matches 675; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 86 ttgcactctgtgacgccccttaccacgaagcttctgtctcttggaagcccttggaattc 145
 |||||||
 Db 1 TTGCTACTTGTGAGCTGCGCCCTTACACAGAGTTTGTTGTTCTGGAAAGCCCTGGATTC 60
 QY 146 gctaaatacctatacactgtgaagtgctgaagggaacagatgaagacatgacctcaaggag 205
 |||||||
 Db 61 GCTAAATACCTATCATCTAGAGTCTGAAGGAAACAGATGAACATGACCTCAAGGAG 120
 QY 206 ctctcctgtaataagagacacagcctgagcctgagcaagaataataaaggagagcctgaa 265
 |||||||
 Db 121 CTTCCTGTCAATGAGAGACCAAGCTGACGCTGGCAAAAGATATTAAGAGAGCTGA 180
 QY 266 actgtctctgtgacatctatagtatgtcagaanaataccttctgtgaaggttagaagatcag 325
 |||||||
 Db 181 ACTGTTCCCTTGAGACATCTTATGATGTCAAGAAATACC--TTTGAGAGGTTAAGAAATCAG 229
 QY 326 gggacatggtgtgtcacattgtctgccacaggaacccgcagctctcacttggaacaga 385
 |||||||
 Db 240 GGGACATGGTGTGTCACTTGTGCGCCACGAAACCGCAGCTTCACTTGAAGACGA 299
 QY 386 atcacgcctgtgaagaagatcatccctaaagcagagagaagctactcaagaagatgtgttc 445
 |||||||
 Db 300 ATCACGCTTGTGAAGAGATCATCTTACACAGAGAGAAAGCTACTTAAGGATTTGTCTC 359
 QY 446 tccctcaaccttccctgtctgcgtgtctccacctgtctccactctgtgaagatgttcaat 505
 |||||||
 Db 360 TCTCCACACTTCCCTGCTGCTGCGGTCCACACTGTCTCCATCTGTGACGATGTTCAAT 419
 QY 506 ggaagagacctctgcagctgcatctactgtgtggtcctgcgtctctatgtctgtgcc 565
 |||||||
 Db 420 GGAAGAGACTCTCCAGCTGCATTAATCTGTGGCTCTGGGCTCTGTATGCTCTCTGG-CA 478
 QY 566 ctgtgctctgaacacttctcctcaggtgtgaaggtgtgactgtgacacactgtggtctgagt 625
 |||||||
 Db 479 CTGTGGCTGTGAACACTTCTTTCAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 538
 QY 626 ccagggaatctcaaacagcagctactgtagaataatctgtataatctcctgaacactccag 685
 |||||||
 Db 539 CCAGGGAATCTCAAAAGCACTACTGTAGGAATATCTGTATATTTCTGAAACCTCCAG 598
 QY 686 caaagaggtctatcaactgtctcaggggtcccgagagggcgaaggagggcaggtctcagg 745
 |||||||
 Db 599 CAAGAAGGTATATCACTGTTCAGGGGTCAACCCGAGAGGACCAAGAGGCAAGTCTTCAGG 658
 QY 746 ctattctgaataacactgtgag 765
 |||||||
 Db 659 CTATTCGATTAACCTGGAG 678

RESULT 15
LOCUS BG328280 955 bp mRNA linear EST 27-FEB-2001
DEFINITION 602427256F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546813 5',
 mRNA sequence.
ACCESSION BG328280
VERSION BG328280.1 GI:13134627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 955)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM1233 row: P column: 14
High quality sequence stop: 718.

FEATURES
Source Location/Qualifiers
1..955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4546813"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(C). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 245 a 230 c 246 g 233 t 1 others
ORIGIN

Query Match 28.1%; Score 650.8; DB 10; Length 955;
Best Local Similarity 96.9%; Pred. No. 3.9e-178;
Matches 717; Conservative 0; Mismatches 17; Indels 6; Gaps 5;

QY 64 ccttaagtcagggaactgacctgcttactgtgacctgaccttactcagcagttttg 123
DB 2 cttctagtcacgsggaactgccc-tgctacttgtagctgacctgaccttactcagcag- 59
QY 124 ttctgggaagccctggagcttctgctaatactacatcacctgaagtgctgaaggaaacaga 183
DB 60 ttctgggaagccctggagcttctgctaatactacatcacctgaagtgctgaaggaaacaga 119
QY 184 tgaagaacatgacactcaagaagcttctgcaatgaagaacccaagcctgacaa 243
DB 120 tgaagaacatgacactcaagaagcttctgcaatgaagaacccaagcctgacaa 179
QY 244 agatatlaagaagagcctgaaactgttctctggacatctatgaatgtcagaanaatacc 303
DB 180 agatatlaagaagagcctgaaactgttctctggacatctatgaatgtcagaanaatacc 239
QY 304 ttlttgagggttaagaagatcaagggaacatggtgttcaacattgtgctgcacggaaacag 363
DB 240 ttlttgagggttaagaagatcaagggaacatggtgttcaacattgtgctgcacggaaacag 299
QY 364 ccagcttcaacttgaaacagaatcacgctgtgaaagatcatccccaagcagagag 423
DB 300 ccagcttcaacttgaaacagaatcacgctgtgaaagatcatccccaagcagagag 359
QY 424 aagctactaagaagatgtgtcctcctcgaactcctgtgtcgtgctcgaactgtctcc 483
DB 360 aagctactaagaagatgtgtcctcctcgaactcctgtgtcgtgctcgaactgtctcc 419
QY 484 catcttgtaagatggttcaatgtaagaagagactgtcgaactgcatctactgtggtgtc 543
DB 420 catcttgtaagatggttcaatgtaagaagagactgtcgaactgcatctactgtggtgtc 479
QY 544 ggcctgataatgctgctgccaactgtggtcctcgaacttctcagttgaagtgtgac 603

DB 480 ggcctgataatgctgctgccaactgtggtcctcgaacttctcagttgaagtgtgac 539
QY 604 tctgaccacttggtctggaagtcacgaagatcacaagccagactgtagaatcttg 663
DB 540 tctgaccacttggtctggaagtcacgaagatcacaagccagactgtagaatcttg 599
QY 664 tataattcctg-aaacttcagcaagaaggtctatcaactgttcagggttccaccgag 722
DB 600 tataatttcttgaaacttccagcaagaaggtctatcaactgttcaagggtcacc--gag 657
QY 723 ggaaccaagaagagtgcttcaagctatctgaataacctggaggtcagaagaagcgaga 782
DB 658 ggaaccaagaagagtgcttcaagctatctgaataacctggaggtc-agaagaagcgaga 716
QY 783 ggccttcacagacaccact 802
DB 717 gctttacagaAACCAATACT 736

Search completed: September 26, 2002, 07:00:01
Job time: 14156 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 01:53:55 : Search time 4629.8 Seconds
(without alignments)
1107.392 Million cell updates/sec

Title: US-09-874-390-1_COPY_1_245

Perfect score: 245

Sequence: 1 attaactgggtttctctatt.....caaagctgacgcctgccaag 245

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: GenEmbl:*
2: gb_hg:*
3: gb_in:*
4: gb_com:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
------------	-------------	-------	--------	-------	-------------

1	245	100.0	174323	2	AC092754	AC092754 Homo sapi
2	245	100.0	187275	2	AC092755	AC092755 Homo sapi
3	176	71.8	2216	6	AX045249	AX045249 Sequence
4	176	71.8	2217	6	AX045251	AX045251 Sequence
5	176	71.8	2217	6	AX045253	AX045253 Sequence
6	58	23.7	2106	9	BC017032	BC017032 Homo sapi
7	37	15.1	101204	9	AC109761	AC109761 Rattus no
8	37	15.1	109101	9	AL136446	AL136446 Human DNA
9	37	15.1	170695	2	AC068545	AC068545 Homo sapi
10	37	15.1	253217	2	AC016590	AC016590 Homo sapi
11	36.8	15.0	156733	9	AC069411	AC069411 Homo sapi
12	36.8	15.0	193295	2	AC073093	AC073093 Homo sapi
13	35.4	14.4	150650	2	AC021770	AC021770 Homo sapi
14	35.4	14.4	161733	2	AC013635	AC013635 Homo sapi
15	35.4	14.4	192696	9	AC009869	AC009869 Homo sapi
16	35	14.3	39882	9	AL356289	AL356289 Human DNA
17	35	14.3	59905	2	AC100985	AC100985 Mus muscu
18	35	14.3	98985	2	AL139427	AL139427 Homo sapi
19	35	14.3	125290	9	AC004986	AC004986 Homo sapi
20	35	14.3	146255	2	AC084170	AC084170 Homo sapi
21	35	14.3	178984	2	AC018540	AC018540 Homo sapi
22	35	14.3	212712	2	AC063940	AC063940 Homo sapi
23	34.8	14.2	199211	2	AC097054	AC097054 Rattus no
24	34.6	14.1	193564	2	AC084779	AC084779 Mus muscu
25	34.6	14.1	235693	2	AC084780	AC084780 Mus muscu
26	34.4	14.0	104115	2	AC022426	AC022426 Homo sapi
27	34.4	14.0	105565	2	AC104103	AC104103 Mus muscu
28	34.4	14.0	117735	9	AC022440	AC022440 Homo sapi
29	34.4	14.0	148103	9	AL445232	AL445232 Human DNA
30	34.4	14.0	201129	2	AC011358	AC011358 Homo sapi
31	34.2	14.0	100834	9	AL160411	AL160411 Human DNA
32	34.2	14.0	251492	2	AL663077	AL663077 Mus muscu
33	34.2	14.0	292494	2	HSJ39370	HSJ39370 Homo sapi
34	33.8	13.8	65243	2	AC104669	AC104669 Homo sapi
35	33.8	13.8	162932	2	AC018516	AC018516 Homo sapi
36	33.6	13.7	97225	2	AC099417	AC099417 Rattus no
37	33.6	13.7	105891	9	AC004850	AC004850 Homo sapi
38	33.6	13.7	121684	2	AL354655	AL354655 Homo sapi
39	33.6	13.7	125787	9	AL645820	AL645820 Human DNA
40	33.6	13.7	125787	9	HSJ13817A	HSJ13817A Human DNA
41	33.6	13.7	172000	1	AF000005	AF000005 Erythrocyt
42	33.6	13.7	197620	9	AF001905	AF001905 Homo sapi
43	33.4	13.6	35822	9	AC004609	AC004609 Homo sapi
44	33.4	13.6	105040	9	AC011229	AC011229 Homo sapi
45	33.4	13.6	148005	2	AC024944	AC024944 Homo sapi

ALIGNMENTS

RESULT 1
AC092754 174323 bp DNA linear HTG 19-JAN-2002
LOCUS Homo sapiens chromosome 15 clone RP11-112N19 map 15q22, ***
DEFINITION SEQUENCING IN PROGRESS ***, 3 ordered pieces.
ACCESSION AC092754
VERSION AC092754.3 GI:18249988
KEYWORDS HTG, HTGS_PHASE2, HTGS_ACTIVERPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 174323)
AUTHORS Rowen, J., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dots, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D., and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174323)
AUTHORS Rowen, J., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dots, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D., and Hood, L.
TITLE Direct Submission

JOURNAL

Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
On Jan 19, 2002 this sequence version replaced gi:15022676.

COMMENT

Genome Center

Center: Multimegabase Sequencing Center

Center code: UMNISC

Web site: http://chroma.mbt.washington.edu/msg_www

Contact: leerowensystemsbiology.org

----- Summary Statistics -----

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-terminator Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

Insert size: -----; agarose-1p

Quality coverage: ---x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 99776: contig of 99776 bp in length

* 99777 99876: gap of unknown length

* 99877 172125: contig of 72249 bp in length

* 172126 172225: gap of unknown length

* 172226 174323: contig of 2098 bp in length.

Location/Qualifiers

1. 174323

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="15"

/map="15q22"

/clone="RP11-112N19"

/clone_lib="RPCT human BAC library 11"

/note="This clone overlaps RP11-99L18 and RP11-361D15"

BASE COUNT 46623 a 36456 c 37881 g 53163 t 200 others

ORIGIN

Query Match 100.0%; Score 245; DB 2; Length 174323;

Best Local Similarity 100.0%; Pred. No. 3.7e-70;

Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attactggtttccattatcatatcctcgcattactcctcgtgtagagccctc 60

|||||

Db 172881 ATTAACTGCGTTTCTATTATCTATCTCGCATTTACTTCTGTAGTAGAGCCCTCT 172940

|||||

QY 61 tctcttaagtaacggaagactcctgtgactgtgaccccttactcaacagctt 120

|||||

Db 172941 TCTCTTAAGTACGAGGAACTGCCCTTGTGACTGTCGCCCTTTATACAGAGTTT 173000

|||||

QY 121 ttgtcttggaagccctcgtggttctgctataactatacctgtagtgctgaaaggaac 180

|||||

Db 173001 TTGTTCTGGGAAGCCCTGGGATTCTGCTAAATCACTATCACTGAGTGTCTGAAGGAAC 173060

|||||

QY 181 agatgaagaacatgacctcaagagagctcgtgctaaatgaagaagccaaagcgcgcgg 240

|||||

Db 173061 AGATGAAGAATCACTCAAGAGCTTCTGTCAATGAGAGCAAGCTGACGCTGG 173120

QY 241 caaag 245

|||||

Db 173121 CAAAG 173125

RESULT 2

AC092755

LOCUS

DEFINITION

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

AC092755

COMMENT

Genome Center

Center: Multimegabase Sequencing Center

Center code: UMNISC

Web site: http://chroma.mbt.washington.edu/msg_www

Contact: leerowensystemsbiology.org

----- Summary Statistics -----

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-terminator Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

Insert size: -----; agarose-1p

Quality coverage: ---x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 32680: contig of 32680 bp in length

* 32681 32780: gap of unknown length

* 32781 187275: contig of 154495 bp in length.

Location/Qualifiers

1. 187275

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="15"

```

/map="15q22"
/clone="RP11-361D15"
/clone_lib="RP11 human BAC library 11"
/note="This clone overlaps RP11-112N19 and RP11-566G20"
BASE COUNT 53668 a 38780 c 38083 g 56624 t 100 others
ORIGIN

```

Query Match	100.0%	Score 245;	DB 2;	Length 187275;
Best Local Similarity	100.0%;	Pred. No. 3.8e-70;		
Matches 245;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 attaaactggatttccctattatctatcctctcgacttaacttctctgagtcagagcctct 60
|||||
Db 33616 ATTAACGTGGGTTTCCATTTATCTATTCCTCTGCCATTACTTCTCTGAGTCAGAGCCCTCT 33675

QY	61	tccctctaagtcacgggaactgccttgccttgtagacctgcccttaactaacgattt	120
Db	33676	TCCTCTAAGTCAACGGGAACGCCCTTGCTACTGTTGACCTGCCCTTTACTCAGACATT	33735

Qy 121 ttgtcttggaagccctggatctctgcaatacctatacattgtagtctgtgaagggaaac 180
|||||
Db 33736 ttgttctggaagccctggattctgcaatacctatcacgtagctctgaagggaac 33795
|||||

Dy **181** agatgagaacatgaactcaaggagcttcctgttcaatgatgaaaccaagtgcgcccg **240**

 |||||
Db 33796 AGATGAGAACATGACTCAAGGAGCTTCCGTTCATGAGAAGACCAAGCTGACGCCCGG 33855

Qy 241 caaag 245
 |||||
 Db 33856 CAAG 33860

RESULT	3			
AX045249				
LOCUS	AX045249	2216 bp	DNA	linear
				PAT 24-NOV-2

REGISTRATION	SEQUENCE 1 FROM FILE# 000000/27.
ACCESSION	AX045249
VERSION	AX045249.1
KEYWORDS	GI:11343799

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 006727-A 1 09-NOV-2000;

FEATURES	Location/Qualifiers
source	1. .2216 /organism="Homo sapiens"

5'UTR	1..426
CDS	427..1743
	/note="unnamed protein product"

```

/protein_id="CAC17334.1"
/db_xref="GI:11343800"
/translation="MWQWRKLQGLHYLMALGCMYLATVALKLSFRPKCDSDHIG
DEGQGVQNTVNETVTDVPSGLVQSCGEMDQDQVQATVNTNKKVQDDEEM

```

YLSLRDCSEHFKAERKFIQPLSKSEVEEPPIASWVTHEKLENFERRLRVAVYAPQ
CVAHDEKSPETFEAAVKALISCPENVFILSKLVRVYVAVSSVQADLNCMEYLQ
PKWYFLNTCGTDFIKSNAEMOVALKMLGRNSMSEVAPSKHKETRMKYHFEEVVR
HITNKKRDPBYNI TMEETGNAY IASPSDVOHV I KND SOOI TEEVVDYSDPEVR

3'UTR	BASE COUNT
1744 . . 2216	600 a 499 c 547 g 570 t

ORIGIN	Query Match	Score	DB	Length
		71.8%	176	2316

Best Local Similarity 100.0%; Pred. No. 2.3e-47;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 70 gtcacgggaactgcccctgtctactgtgacactgccccttaactcagcagltttgtctcgtg 129
 |||||
 b 1 gtcacgggaactgcccctgtctactgtgacactgccccttaactcagcagltttgtctcgtg 60

130 gaagccctggaattctcgtaacactacacgttagtctgtaggaggaacagatgaaga 189
|||||
61 GAAGCCCTGGATTCTGCTAAACCTATCAGTGTAGGTGCTGAAGGGAAACAGATGAAGA 120

190 acatgacctcaagagagcttcctgctcaatgagaagaccagctgacgctgcaaaag 245
121 ACATGACCTCAAGGAGCTTCCCTGCTCAATGAGAAGACCAAGCTGACGCCCTGGCAAAAG 176

RESULT	4
DOCUS	
X0A5251	
AX045251	
2217 bp	
DNA	
linear	
PAT 24-NOV-2000	

KEYWORDS
SEQUENCE 3 FROM PATENT WO0006021.
AX045251
AX045251.1 GI:11343801

ORGANISM Homo sapiens
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

ADOLF, G., HELDER, K.H. and SOMMERGRUBER, W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 3 09-NOV-2000;

FEATURES	Location/Qualifiers
source	1. .2217
	/organism="Homo sapiens"

5'UTR	1. .844
5'UTR	1. .426
CDS	427. .999

```

/codon_start=1
/protein_id="CAC17335.1"
/db_xref="GI:11343802"
/tranlation="MVKKPKDQYUVAIGCVATATVAIKTSEDKKQSDNHTGIES

```

RESOSQYCRNLLYFLFLPAKMSINCSTGEOEAVLOAILNNEVKKRREPETH
YLSLTPDCEHEKAEKRFIQPLSKEEVEFPIAVSMVIERKIEENRLLRAVYAPONIY
CVHDEEVPNRNFQGGQSNYFLFLPKCLHSQ" 845 1744

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17336.1"
/db_xref="GI:1343803"

```

```

translation="MRKRKILGKCELDMLRITVYSMMNKSPTFKKAVKALISCF
/PNVTASKLVRVYVASWSRVQADLNCMDLQSSVPWKVFKFLNTGSDTFLKSNAEYV
ALKMLNGRSMSESEVPRKHETRNKHYEEVVRDLHLTKNNKDDPPYNLMTGNAYI
VASRDEVOHVLKNPKSOQLIEWKDKTYSDEHLMATLOARMMSPSVNPNKPKYDSDM

```

3'UTR
1000..2217
3'UTR
1745..2217

REG. COMM.	500	400	300	200
RIGIN	500	400	300	200

Best Local Similarity 100.0%; Pred. No. 2.3e-47;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTCACGGGAACTGCCCTTGCTACTTGACCTGCCCTTACTACGACATTTTGTCTCG 60

[illegible]

```

Db      61 GAAGCCCTGGGATTCTGCTAATACCTATCACTAGTGTGGAAGGGAACAGATGAAGA 120
        |||||||
Qy      190 acatgaacctaaagagcttcctgtaaatgagaagaccactgaacgtggaag 245
        |||||||
Db      121 ACATGACCTCAAGAGCTTCTCTCAATGAGAAACCAAGCTGACGCTGCAAG 176

RESULT  5
AX045253
LOCUS   AX045253                2217 bp    DNA          linear    PAT 24-NOV-2000
DEFINITION   Sequence 5 from Patent WO006727.
ACCESSION   AX045253
VERSION     AX045253.1   GI:11343804
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 2217)
AUTHORS    Adolf,G., Heider,K.H. and Sommergruber,W.
TITLE       Tumour-associated antigen
JOURNAL     Patent: WO 006727-A 5 09-NOV-2000;
            Boehringer Ingelheim International GmbH (DE)
FEATURES
    source
        1..2217
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /rname="1"
            /protein_id="C11343805"
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="C11343805"
            /translation="MRLKTLKGYCELCMPRLTYTVSMKKSPETFEPAVALISCF
            PVFTASKIVRYVYASMSRVQADLNCMEDLQSSVPRKYLINCCGDPRIKSAEWQ
            ALKMLNGRSMSEVPKPKHETRMKTHFEVYRDTLHLNKKDPPRYNLTFTGNAYI
            VASRDVQVHLNPNKSQQLIEWKDYSPDEHLMATLQARWMPGSPVPHRYDIDSM
            TSIRARLVKQWQHGIDDKGAPAPCSGIHQRAICVYGAGDLNMLNHLNKKFDPK
            VDDNALQCLEEYLRKAYITEL"
            1745..2217

3'UTR    600 a 498 c 547 g 572 t
BASE COUNT
ORIGIN

Query Match      71.8%; Score 176; DB 6; Length 2217;
Best Local Similarity 100.0%; Pred. No. 2.3e-47;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      70 gtaacgggaactgccttctactgtgacctgccttactaagaagtttctctg 129
        |||||||
Db      1 GTCACGGGAACCTCCCTGCTACTGTGACCTGCCCTTACTCAGCAGTTTGTCTG 60

Qy      130 gaagccctggagattctgtaatacctatacctctgtagtctgaagggaacagatgaaga 189
        |||||||
Db      61 GAAGCCCTGGGATTCTGCTAATACCTATCACTAGTGTGGAAGGGAACAGATGAAGA 120

Qy      190 acatgaacctaaagagcttcctgtaaatgagaagaccactgaacgtggaag 245
        |||||||
Db      121 ACATGACCTCAAGAGCTTCTCTCAATGAGAAACCAAGCTGACGCTGCAAG 176

RESULT  6
BC017032
LOCUS   BC017032                2106 bp    mRNA          linear    PRI 09-NOV-2001
DEFINITION   Homo sapiens, glucosaminyl (N-acetyl) transferase 3, mucin type,
ACCESSION   BC017032
VERSION     BC017032.1   GI:16877561
KEYWORDS    MGC.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE   1 (bases 1 to 2106)
AUTHORS    Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (05-NOV-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www.shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxli.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: TRAK Plate: 20 Row: n Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4758421.

FEATURES
    source
        1..2106
            /organism="Homo sapiens"
            /db_xref="LocusID:9245"
            /db_xref="taxon:9606"
            /clone="MGC:9086 IMAGE:3851937"
            /tissue_type="Colon, adenocarcinoma"
            /clone_lib="NIH_MGC_65"
            /lab_host="DH10B"
            /note="vector: pCMV-SPORT6"
            309..1625
            /product="glucosaminyl (N-acetyl) transferase 3, mucin
            type"
            /protein_id="AAH17032.1"
            /db_xref="GI:16877562"
            /translation="MYQWRKLQSLHLYALGCMILATYALKLSPRLKDSPHLGIES
            RESOYOYCRNLILYNELKLPKRSINSGVTRGDQEAVALIINLEEVKKRPFYDTH
            YLSLRDCEHFAERKFIOPLSKSEVEPIAYSWIHRIENFERLAAVAPQNTY
            CVHDEKSPETFEKAVKAIISCFPNVFIASKLAVRYVYASMSRVQADLNCMEDLQSSV
            PKRYFNLTGTPDKPIKSNAMQALKMLNGRSMSEVPKPKHETRMKTHFEVYRDTL
            HLNKKKDDPPRYNLTFTGNAYIVASRDVQVHLNPNKSQQLIEWKDYSPDEHLMAT
            TLQARWMPGSPVPHRYDIDSMTSIRARLVKQWQHGIDDKGAPAPCSGIHQRAICV
            YGAGDLNMLNHLNKKFDPKVDNALQCLEEYLRKAYITEL"
            471 c 516 g 537 t

BASE COUNT 582 a 471 c 516 g 537 t
ORIGIN

Query Match      23.7%; Score 58; DB 9; Length 2106;
Best Local Similarity 100.0%; Pred. No. 3.7e-08;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      188 gaacatgacctaaagagcttcctgtaaatgagaagaccactgaacgtggaag 245
        |||||||
Db      1 GAACATGACCTCAAGAGCTTCTCTCAATGAGAAACCAAGCTGACGCTGCAAG 58

RESULT  7
AC109761
LOCUS   AC109761                101204 bp    DNA          linear    HTG 07-FEB-2002
DEFINITION   Rattus norvegicus clone CH230-310K9, *** SEQUENCING IN PROGRESS
ACCESSION   AC109761
VERSION     AC109761.1   GI:18580874
KEYWORDS    HTG: HTGS_PHASE1.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus

```

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 101204)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshan, F.R., Allen, C.,
Albrooks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T.,
Barbata, J., Benton, J., Blum, K., Blankenburg, K., Bonini, D.,
Bouck, J., Boyle, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Daya, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
Falls, J., Ferraguto, D., Flagg, N., Ford, U., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kritov, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhley, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Mosier, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okunolu, G.,
Orduy, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Syatke, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tatney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watson, R., Wall, R.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
Direct Submission
2 (bases 1 to 101204)

Worley, K.C.

Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GGV1

Center clone name: CH230-310K9

Summary Statistics

Sequencing vector: Plasmid; M77789

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 82934 bases at least Q40

Consensus quality: 87376 bases at least Q30

Consensus quality: 90911 bases at least Q20

Estimated insert size: 83906; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 1.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	5456:	contig of 5456 bp in length
5457	3556:	gap of unknown length
5557	10707:	contig of 5151 bp in length
10708	10807:	gap of unknown length
10808	16279:	contig of 5472 bp in length
16280	16379:	gap of unknown length
16380	21128:	contig of 4749 bp in length
21129	21228:	gap of unknown length
21229	24409:	contig of 3181 bp in length
24410	24509:	gap of unknown length
24510	27744:	contig of 3235 bp in length
27745	27844:	gap of unknown length
27845	30710:	contig of 2866 bp in length
30711	30810:	gap of unknown length
30811	34857:	contig of 4047 bp in length
34858	34957:	gap of unknown length
34958	37581:	contig of 2624 bp in length
37582	37681:	gap of unknown length
37682	40588:	contig of 2907 bp in length
40589	40688:	gap of unknown length
40689	43200:	contig of 2512 bp in length
43201	43300:	gap of unknown length
43301	47856:	contig of 4556 bp in length
47857	47956:	gap of unknown length
47957	50293:	contig of 2337 bp in length
50294	50393:	gap of unknown length
50394	52992:	contig of 2599 bp in length
52993	53092:	gap of unknown length
53093	56555:	contig of 3463 bp in length
56556	56655:	gap of unknown length
56656	59193:	contig of 2588 bp in length
59194	59293:	gap of unknown length
59294	61403:	contig of 2110 bp in length
61404	61503:	gap of unknown length
61504	63792:	contig of 2289 bp in length
63793	63892:	gap of unknown length
63893	65018:	contig of 1126 bp in length
65019	65118:	gap of unknown length
65119	67500:	contig of 2382 bp in length
67501	67600:	gap of unknown length
67601	68960:	contig of 1360 bp in length
68961	69060:	gap of unknown length
69061	71414:	contig of 2354 bp in length
71415	71514:	gap of unknown length
71515	73216:	contig of 1702 bp in length
73217	73316:	gap of unknown length
73317	74830:	contig of 1514 bp in length
74831	74930:	gap of unknown length
74931	76270:	contig of 1340 bp in length
76271	76370:	gap of unknown length
76371	77794:	contig of 1424 bp in length
77795	77894:	gap of unknown length
77895	79050:	contig of 1156 bp in length
79051	79150:	gap of unknown length
79151	80610:	contig of 1460 bp in length
80611	80710:	gap of unknown length
80711	81875:	contig of 1165 bp in length
81876	81975:	gap of unknown length
81976	83545:	contig of 1570 bp in length
83546	83645:	gap of unknown length
83646	85044:	contig of 1399 bp in length
85045	85144:	gap of unknown length
85145	86842:	contig of 1658 bp in length
86843	86942:	gap of unknown length
86943	88741:	contig of 1799 bp in length


```

repeat_region /note="THE1B-INTERNAL repeat: matches 1. .433 of consensus"
20697. .21060
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 21164. .21371
/note="MIR repeat: matches 34. .251 of consensus"
21581. .21916
/note="Alu10 repeat: matches 1. .312 of consensus"
24571. .24662
/note="L2 repeat: matches 2334. .2428 of consensus"
repeat_region 26124. .26422
/note="AluY repeat: matches 1. .299 of consensus"
28540. .28740
/note="L2 repeat: matches 1535. .1713 of consensus"
repeat_region 28741. .29040
/note="Alu10 repeat: matches 1. .297 of consensus"
29041. .29493
/note="L2 repeat: matches 990. .1535 of consensus"
29493. .29999
/note="L2 repeat: matches 1309. .1860 of consensus"
32968. .33150
/note="MIR repeat: matches 65. .262 of consensus"
35906. .36171
/note="LIME repeat: matches 937. .1211 of consensus"
36608. .38632
/note="L2 repeat: matches 202. .2471 of consensus"
38708. .38899
/note="2 copies 96 mer 80% conserved"
39435. .39762
/note="MER58B repeat: matches 1. .340 of consensus"
40470. .40495
/note="13 copies 2 mer aa 92% conserved"
41247. .41326
/note="MAD1 repeat: matches 1. .80 of consensus"
43677. .43997
/note="L2 repeat: matches 2402. .2708 of consensus"
48911. .49064
/note="L1MB1 repeat: matches 5628. .5776 of consensus"
49065. .49311
/note="L1PA13 repeat: matches 5909. .6155 of consensus"
49313. .49684
/note="L1PA10 repeat: matches 5774. .6158 of consensus"
50119. .50342
/note="L1MB1 repeat: matches 5935. .6170 of consensus"
51081. .51267
/note="MIR repeat: matches 46. .245 of consensus"
51273. .51316
/note="L1R25 repeat: matches 145. .190 of consensus"
52873. .52915
/note="MIR repeat: matches 211. .253 of consensus"
52874. .52926
/note="L2 repeat: matches 2694. .2746 of consensus"
56033. .56405
/note="MST1 repeat: matches 1. .394 of consensus"
56465. .56555
/note="L2 repeat: matches 2579. .2671 of consensus"
56556. .56875
/note="MER2 repeat: matches 1. .345 of consensus"
56876. .57207
/note="L2 repeat: matches 2262. .2579 of consensus"
58300. .58567
/note="AluY repeat: matches 24. .303 of consensus"
60945. .61032
/note="MIR repeat: matches 61. .145 of consensus"
61490. .61596
/note="L2 repeat: matches 2577. .2691 of consensus"
62253. .62477
/note="MIR repeat: matches 20. .252 of consensus"
62484. .63537
/note="MER10 repeat: matches 1. .1071 of consensus"
63966. .65052
/note="L1MB8 repeat: matches 5088. .6173 of consensus"
68422. .69775
/note="Charliela repeat: matches 7. .1455 of consensus"

```

```

repeat_region 70089. .70150
/note="31 copies 2 mer at 79% conserved"
repeat_region 70091. .70150
/note="15 copies 4 mer atat 80% conserved"
misc-feature 71822. .72082
/note="Sequence from overlapping clone AU356425. Assembly
confirmed by restriction digests."
repeat_region 72162. .72468
/note="AluY repeat: matches 1. .305 of consensus"
72866. .73038
/note="MER5A repeat: matches 7. .189 of consensus"
73289. .73615
/note="Alu5x repeat: matches 1. .302 of consensus"
74557. .74623
/note="MER47 repeat: matches 1. .66 of consensus"
74977. .75018
/note="21 copies 2 mer aa 76% conserved"
75398. .75700
/note="Alu5x repeat: matches 1. .305 of consensus"
76142. .76605
/note="L1ME2 repeat: matches 4902. .5347 of consensus"
76606. .76911
/note="Alu1b repeat: matches 1. .312 of consensus"
76912. .77672
/note="L1ME2 repeat: matches 5347. .6155 of consensus"
77727. .77781
/note="MER5A repeat: matches 111. .167 of consensus"

Query Match 15.1%; Score 37; DB 9; Length 109101;
Best Local Similarity 57.3%; Pred. No. 0.47;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

```

QY 6 ctgggtttccattatatacctctgcatactctctgaagcagcctctct 65
Db 17946 CTCCTTCTCTATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17887
QY 66 ctcaagtcacgaagactgcctctactctgtacactgccttactcaagatttt 122
Db 17886 TTCTGCTGAGAGACTGCCCTTCCTGCTTTCGATTTTCACACCTCCACATTCTCT 17850

RESULT 9
AC068545 170695 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-385M12, WORKING DRAFT
DEFINITION AC068545
ACCESSION AC068545.3 GI:8099098
VERSION AC068545.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170695)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL unpublished
REFERENCE 2 (bases 1 to 170695)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On May 26, 2000 this sequence version replaced gi:7820143.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0385M12
----- Summary Statistics -----
Sequencing vector: M13; 100%

```



```

* 31908 32007: gap of unknown length
* 32008 36364: contig of 4357 bp in length
* 36365 36464: gap of unknown length
* 36465 43309: contig of 6845 bp in length
* 43310 43409: gap of unknown length
* 43410 48185: contig of 4776 bp in length
* 48186 48285: gap of unknown length
* 48286 55626: contig of 7341 bp in length
* 55627 55726: gap of unknown length
* 55727 65781: contig of 10055 bp in length
* 65782 65881: gap of unknown length
* 65882 69673: contig of 3792 bp in length
* 69674 69773: gap of unknown length
* 69774 79389: contig of 9616 bp in length
* 79390 79489: gap of unknown length
* 79490 91453: contig of 11964 bp in length
* 91454 91553: gap of unknown length
* 91554 96671: contig of 5118 bp in length
* 96672 96771: gap of unknown length
* 96772 100544: contig of 3773 bp in length
* 100545 100644: gap of unknown length
* 100645 111423: contig of 10779 bp in length
* 111424 111523: gap of unknown length
* 111524 115278: contig of 3755 bp in length
* 115279 124011: contig of 8633 bp in length
* 124012 124111: gap of unknown length
* 124112 138475: contig of 14364 bp in length
* 138476 138575: gap of unknown length
* 138576 140613: contig of 2038 bp in length
* 140614 140713: gap of unknown length
* 140714 144949: contig of 4236 bp in length
* 144950 145049: gap of unknown length
* 145050 164352: contig of 19303 bp in length
* 164353 164452: gap of unknown length
* 164453 170540: contig of 6088 bp in length
* 170541 170640: gap of unknown length
* 170641 185046: contig of 14406 bp in length
* 185047 185146: gap of unknown length
* 185147 222309: contig of 37063 bp in length
* 222310 223217: gap of unknown length
* 223217 253217: contig of 30908 bp in length.

```

FEATURES

Location/Qualifiers

1..253217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-3220F14"
/clone.lib="Caltech human BAC library D"

BASE COUNT 62246 a 62867 c 62532 g 61752 t 3820 others
ORIGIN

Query Match 15.1%; Score 37; DB 2; Length 253217;
Best Local Similarity 64.7%; Pred. No. 0.5;

Matches 55; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 11 ttcttcattctatctatctcgcattctctctcgcagtcagagccttctcttaag 70
Db 187609 ttttctctatgtatgtatctcttattttatttattatgagacagatctgcctctgtctac 187550

QY 71 tcacgggaactgcctctgactctg 95
Db 187549 acgagagatgcgcacatgccatgtg 187525

RESULT 11

AC069411

LOCUS AC069411 156733 bp DNA linear PRI 27-SEP-2001
DEFINITION Homo sapiens 3 BAC RPL1-609K13 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.

ACCESSION AC069411
VERSION AC069411.21 GI:13876417

KEYWORDS

HTG.
human.

SOURCE
ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 156733)
Muzny D.M., Adams C.C., Adio-Oduola B., All-ouman F.R., Allen C.,
Alsdorfs S.L., Amarantunga H.C., Are J.R., Banks T., Barbara J.,
Benton J., Blum K., Blankenburg K., Bonnin D., Bouck J.,
Bowle S., Brileva M., Brown E., Brown M., Bryant N.P., Bunay C.,
Burich P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F.,
Carter M., Cavazos S.R., Chacko J., Chavez D., Chen R.,
Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,
Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
Day-Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
Denn A.L., Ding F., Dinh H.H., Douthwaite K.J., Draper H.,
Dugan-Hochia S., Durbin K.J., Earmhart C., Edgar D., Edwards C.C.,
Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,
Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,
Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,
Hamilton K., Harris C., Harris K., Hart M., Havlik P., Hawes A.,
He X., Hernandez J., Hernandez O., Hodgson A., Hognes M.,
Holloway C., Hollins B., Homs J.F., Howard S., Huber J., Huik S.,
Hume J., Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolivet S.,
Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah D.,
Kovar C., Kratovic J., Kuresh A., Landry N., Leal B., Lewis L.C.,
Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W.,
Luissegged H., Lorado R.J., Lu X., Lucier A., Lucier R., Luna R.,
Ma J., Maheshwari M., Mapa P., Martin R., Martindale A.,
Martinez E., Massey E., Mawhney E., McLeod M.P., Meador M.,
Mel G., Metzger M., Miner G., Miner Z., Mitchell T., Mohabbat K.,
Moore S., Morgan M., Moorish T., Morris S., Moser M., Neal D.,
Nelson D., Newton J., Newton N., Nguyen N., Nguyen N.,
Nickerson E., Nockenkw S., Oguh B., Okunou G., Oragunye N.,
Oviedo R., Pace A., Payton B., Peery J., Perez L., Peters L.,
Pickens R., Primus E., Pu L.L., Quiles M., Ren Y., Rives M.,
Rojas A., Rojubokan I., Rolfe M., Ruiz S., Savery G., Scherer S.,
Scott G., Shen H., Shoshitari N., Sison I., Sodergren E.,
Sonalke T., Sparks A., Stanley H., Stone H., Sutton A., Svatek A.,
Tabors P., Tamerisa A., Tamerisa K., Tang H., Tansey J., Taylor C.,
Taylor T., Tellrod B., Thomas N., Thomas S., Usmani K., Vasquez L.,
Vera V., Villalón D., Vinson R., Wall R., Wang S., Usmani K., Vasquez L.,
Warren R., Washington C., Watlington C., Williams G.,
Williamson A., Wleczek R., Woodson S., Worley K., Wu C.,
Wu Y.F., Zhou J., Zorilla S., Naylor S.L. and Gibbs R.

TITLE

Journal Submission

JOURNAL

Unpublished
2 (bases 1 to 156733)

REFERENCE

Worley K.C.
Direct Submission
Submitted (30-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

Worley K.C.
Direct Submission
Submitted (28-APR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

Worley K.C.
Direct Submission
Submitted (01-MAY-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

Worley K.C.
Direct Submission
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

Worley K.C.
Direct Submission
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

Worley K.C.
Direct Submission
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

Worley K.C.
Direct Submission
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Apr 28, 2001 this sequence version replaced gi:13811719.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-helpebcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

Location/Qualifiers

source

```
1. .156733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-609K13"
1. .147
/rpt_family="L1M2"
repeat_region
218. .428
/rpt_family="L1M2"
repeat_region
691. .1451
/rpt_family="L1M2"
repeat_region
1481. .3870
/rpt_family="L1M4A"
repeat_region
5874. .6212
/rpt_family="L1M1"
repeat_region
6209. .6520
/rpt_family="L1M4"
repeat_region
6546. .6580
/rpt_family="AT_rich"
repeat_region
7256. .7277
/rpt_family="AT_rich"
repeat_region
7543. .7626
/rpt_family="MIR"
repeat_region
8402. .8442
/rpt_family="GA-rich"
repeat_region
complement(8762. .8985)
/rpt_family="MER20"
repeat_region
complement(10497. .10671)
/rpt_family="MER117"
STS
10795. .11074
/standard_name="G54755"
/db_xref="dbSTS:83999"
repeat_region
11455. .11758
/rpt_family="AluJo"
repeat_region
12083. .12104
```

```
repeat_region /rpt_family="AT_rich"
14223. .14244
repeat_region /rpt_family="AT_rich"
14275. .14357
repeat_region /rpt_family="CT-rich"
complement(14363. .14441)
/rpt_family="L2"
repeat_region /rpt_family="CT-rich"
17006. .17091
repeat_region /rpt_family="CT-rich"
17100. .17190
repeat_region /rpt_family="TTCC)n"
17199. .17237
repeat_region /rpt_family="T-rich"
complement(17238. .17402)
/rpt_family="AluJ"
repeat_region /rpt_family="T-rich"
17403. .17419
repeat_region /rpt_family="T-rich"
complement(17420. .23556)
/rpt_family="L1PA3"
17755. .17886
STS /standard_name="G19948"
/db_xref="dbSTS:32826"
complement(23564. .23691)
repeat_region /rpt_family="AluJ"
23694. .23796
repeat_region /rpt_family="L2"
complement(23979. .24182)
/rpt_family="MER20"
repeat_region complement(24300. .24457)
/rpt_family="L2"
repeat_region complement(24541. .25128)
/rpt_family="L2"
repeat_region /rpt_family="MER74A"
26769. .27238
repeat_region /rpt_family="MER74A"
27240. .27749
repeat_region /rpt_family="MLT2B"
27750. .27830
repeat_region /rpt_family="MER74A"
complement(28216. .28522)
/rpt_family="Tigger4(Comb1)"
repeat_region complement(28538. .29878)
/rpt_family="Tigger4(Comb1)"
repeat_region complement(29873. .30262)
/rpt_family="Tigger4(Comb1)"
30219. .30269
repeat_region /rpt_family="MER46A"
31408. .31531
repeat_region /rpt_family="MER5B"
31534. .31570
```

Query Match 15.0%; Score 36.8; DB 9; Length 156733;

Best Local Similarity 49.0%; Pred. No. 0.56;

Matches 98; Conservativity 0; Mismatches 102; Indels 0; Gaps 0;

```
QY 2 ttaacggagtttcctattatcaccctcgcattactctccctgagtcagcctctt 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34147 TTATATGGCTTTGACATTTGGTAAGTAGTATGTTCCCTGATGGAATGATCT 34206
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 ctctcgaagtcacgggaacgcctctgctactctgtagccctccttactcaagcttt 121
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34207 TTATTTGGCGTATATCTCTGAGTTCTTGCTCTCTCTCTCTCTCTCTCTCT 34266
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 122 tgtctcggaagccctggagatctctgtaataactataactctgtagtgcgtaaggaaca 181
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34267 ATGATTTGGCAGCACTTGTCAGTCAGATCTAAGTGTAGTGTGAGACATGCTCCT 34326
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 gatgaagaacatgacctca 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34327 GCTTAAAGCCTTATTACAA 34346
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 12
AC073093

LOCUS	AC073093	19395 bp	DNA	linear	HTG 01-SEP-2000
DEFINITION	Homo sapiens chromosome 3 clone RP11-357G12, WORKING DRAFT				
SEQUENCE	18 unordered pieces.				
ACCESSION	AC073093				
VERSION	AC073093.1	GI:834806			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
TITLE	1 (bases 1 to 193295)				
JOURNAL	Waterston,R.H.				
REFERENCE	The sequence of Homo sapiens clone				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 193295)				
JOURNAL	Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (08-JUN-2000) Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
JOURNAL	MO 63108, USA				
COMMENT	<p>----- Genome Center -----</p> <p>Center: Washington University Genome Sequencing Center</p> <p>Center code: WUGSC</p> <p>Web site: http://genome.wustl.edu/gsc/index.shtml</p> <p>----- Project Information -----</p> <p>Center project name: H.NH0357G12</p> <p>----- Summary Statistics -----</p> <p>Sequencing vector: M13; 100%</p> <p>Sequencing vector: plasmid; 0%</p> <p>Chemistry: dye-primer ET; 100% of reads</p> <p>Chemistry: dye-terminator Big Dye; 0% of reads</p> <p>Assembly program: Phrap; version 0.990319</p> <p>Consensus quality: 185033 bases at least Q40</p> <p>Consensus quality: 187477 bases at least Q30</p> <p>Consensus quality: 189141 bases at least Q20</p> <p>Insert size: 18000; agarose-fp</p> <p>Insert size: 19159; sum-of-contigs</p> <p>Quality coverage: 4.94 in Q20 bases; agarose-fp</p> <p>Quality coverage: 4.67 in Q20 bases; sum-of-contigs</p> <p>-----</p> <p>* NOTE: This is a 'working draft' sequence. It currently</p> <p>* consists of 18 contigs. The true order of the pieces</p> <p>* is not known and their order in this sequence record is</p> <p>* arbitrary. Gaps between the contigs are represented as</p> <p>* runs of N, but the exact sizes of the gaps are unknown.</p> <p>* This record will be updated with the finished sequence</p> <p>* as soon as it is available and the accession number will</p> <p>* be preserved.</p> <p>* 1</p> <p>* 1161: contig of 1161 bp in length</p> <p>* 1161: gap of unknown length</p> <p>* 1262: contig of 1346 bp in length</p> <p>* 1262: gap of unknown length</p> <p>* 2608: contig of 1060 bp in length</p> <p>* 2708: contig of 1060 bp in length</p> <p>* 3768: gap of unknown length</p> <p>* 3868: contig of 2347 bp in length</p> <p>* 6215: gap of unknown length</p> <p>* 6315: contig of 2852 bp in length</p> <p>* 9167: gap of unknown length</p> <p>* 9267: contig of 4431 bp in length</p> <p>* 13698: gap of unknown length</p> <p>* 13798: contig of 5931 bp in length</p> <p>* 19729: gap of unknown length</p> <p>* 19829: contig of 5574 bp in length</p> <p>* 25403: gap of unknown length</p> <p>* 25502: contig of 7634 bp in length</p> <p>* 33137: gap of unknown length</p> <p>* 33236: contig of 7019 bp in length</p> <p>* 40256: gap of unknown length</p> <p>* 40356: contig of 10173 bp in length</p> <p>* 50529: gap of unknown length</p> <p>* 50629: contig of 15746 bp in length</p> <p>* 66374: gap of unknown length</p>				

	*	66475	80340:	contig of 13866 bp in length
	*	80341	80440:	gap of unknown length
	*	80441	95176:	contig of 14736 bp in length
	*	95177	95276:	gap of unknown length
	*	95277	112764:	contig of 17488 bp in length
	*	112765	112864:	gap of unknown length
	*	112865	133481:	contig of 20617 bp in length
	*	133482	133581:	gap of unknown length
	*	133582	154215:	contig of 20634 bp in length
	*	154216	154315:	gap of unknown length
	*	154316	193295:	contig of 38980 bp in length.
FEATURES				
Source				
		1.	193295	Location/Qualifiers
			/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/chromosome="3"	
			/clone="RP11-357G12"	
		1.	1161	
misc_feature		/note="assembly_name:Contig10"		
misc_feature		1262 . 2607		
misc_feature		/note="assembly_name:Contig12"		
misc_feature		2708 . 3767		
misc_feature		/note="assembly_name:Contig13"		
misc_feature		3868 . 6214		
misc_feature		/note="assembly_name:Contig14"		
misc_feature		6315 . 9166		
misc_feature		/note="assembly_name:Contig15"		
misc_feature		9267 . 13697		
misc_feature		/note="assembly_name:Contig16"		
misc_feature		13798 . 19728		
misc_feature		/note="assembly_name:Contig17"		
misc_feature		19829 . 25402		
misc_feature		/note="assembly_name:Contig18"		
misc_feature		25503 . 33136		
misc_feature		/note="assembly_name:Contig19		
misc_feature		clone_end:r7		
misc_feature		vector_side:right"		
misc_feature		33237 . 40255		
misc_feature		/note="assembly_name:Contig20"		
misc_feature		40356 . 50528		
misc_feature		/note="assembly_name:Contig21"		
misc_feature		50629 . 66374		
misc_feature		/note="assembly_name:Contig22"		
misc_feature		66475 . 80340		
misc_feature		/note="assembly_name:Contig23"		
misc_feature		80441 . 95176		
misc_feature		/note="assembly_name:Contig24		
misc_feature		clone_end:SP6		
misc_feature		vector_side:left"		
misc_feature		95277 . 112764		
misc_feature		/note="assembly_name:Contig25"		
misc_feature		112865 . 133481		
misc_feature		/note="assembly_name:Contig26"		
misc_feature		133582 . 154215		
misc_feature		/note="assembly_name:Contig27"		
misc_feature		154316 . 193295		
misc_feature		/note="assembly_name:Contig28"		
BASE COUNT	60342 a	36033 c	35646 g	59573 t 1701 others
ORIGIN				
Query Match		15.0%;	Score 36.8;	DB 2; Length 193295;
Best Local Similarity		49.0%;	Pred. No. 0.57;	
Matches	98; Conservative	0;	Mismatches 102;	Indels 0; Gaps 0;
OY	2	ttaactgggtttctcattatctacccctcgcgcatcctctctcgagtcagagccctt 61		
Db	60523	TTAATGGCTTTTGACATTGGTAAAGTAGTATGTTCCTCCCTCGATGGAATAATCATCT 60582		
OY	62	cctctagtgaccgggaacctgccttgctactgtgagcactgcgcccttatcagaagttt 121		
Db	60583	TTAATGGCTAATATCTGAGGTTCTTGCTCCCTCCATCTGGCATTTCGTATGCATTAT 60642		

	Matches	78;	Conservative	0;	Mismatches	71;	Indels	0;	Gaps	0;
Oy	42	tcctgagtcagagcctctctctcctaagtcagggaaactgacctgacttgacct	101							
Db	51904	tcctttatcacagcattgacattttccctcagagaaactctgacctttacatgtgaatt	51963							
Oy	102	gcccttactcagcagtttctgctggaagccctggatctgctaatactacact	161							
Db	51964	tcagtttgctgaaacttttttttttttgccacacattggcccatggaatttctcttact	52023							
Oy	162	gtaggctgctgaaaggaacagatgagaa	190							
Db	52024	ttctaggagagcccgaaatgcatttagaa	52052							

Search completed: September 26, 2002, 04:22:05
Job time: 8890 sec

THIS PAGE BLANK (USPTO)

Human schizophrin
Human immune/haematin
Human immune/haematin
Human immune/haematin
DNA encoding anti-1
DNA encoding anti-1
DNA encoding anti-1
DNA encoding anti-1
DNA encoding anti-1
Human immune/haematin
Human immune/haematin
Human immune/haematin
DNA encoding anti-1
Human nervous system
Soybean tyrosine t
Human nervous system
Human reproductive system
Human PC9-1 genome
Arabidopsis thaliana
Human ORF ORF22510
Human reproductive system
Human immune/haematin
Arabidopsis thaliana
DNA encoding novel
Human polynucleotide
Human cDNA sequence
Human cDNA sequence
Human RNA helicase
Human sperm-specific
Human polynucleotide
Human colon cancer
Human ORF ORF2974
cDNA encoding a human
Human polynucleotide
Arabidopsis melanog

PR 04-DEC-1998; 98DK-0001605.
XX
XX (CLAU/) CLAUSEN H.
XX
XX Clausen H, Schwientek T;
PI
XX WPI: 2000-423407/36.
DR P-PSDB: AAY94492.
XX
XX New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
PT probe for the detection of specified glucosaminyltransferase from
PT other species and related organisms
XX
XX Claim 5; Fig 2; 47pp; English.
XX
XX The present sequence encodes human UDP-N-acetylglucosamine:
CC N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
CC (C2/4Gnt). The protein is the third member of the family of O-glycan
CC beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
CC based O-glycans on to oligosaccharides, glycoproteins and
CC glycosphingolipids. C2/4Gnt can therefore be used in the production of
CC appropriately glycosylated glycoconjugates with particular enzymatic,
CC immunogenic, or other biological or physical properties. The nucleotide
CC sequence is useful as a probe for the detection of C2/4Gnt from other
CC species and related organisms and for the recombinant production of
CC C2/4Gnt polypeptide. The nucleotide sequence was identified by analysis
CC of EST database sequence information. Oligonucleotides derived from EST
CC clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
CC from a human foreskin genomic pl library by 5' RACE PCR. RT-PCR was
CC performed using Colo205 human cell line mRNA in order to produce cDNA
CC for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
CC has been implicated in tumour progression and metastasis.
XX
XX Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other;
SQ
Query Match 100.0%; Score 245; DB 21; Length 2319;
Best Local Similarity 100.0%; Pred. No. 7.4e-74;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attactgggtttccattatcatcctcctgcattacttctctgaagcagctct 60
Db 1 attactgggtttccattatcatcctcctgcattacttctctgaagcagctct 60
QY 61 tctctcaagtcaggggaactgcctctgctactgtgactgaccttaactcagcagtt 120
Db 61 tctctcaagtcaggggaactgcctctgctactgtgactgaccttaactcagcagtt 120
QY 121 ttgttcgggaagccctgggattctgttaactactcactgtgagtgccgaaggaac 180
Db 121 ttgttcgggaagccctgggattctgttaactactcactgtgagtgccgaaggaac 180
QY 181 agatgaagacatgacctcaagaagctcctgtcaatgaagaagcagctgaagcttg 240
Db 181 agatgaagacatgacctcaagaagctcctgtcaatgaagaagcagctgaagcttg 240
QY 241 caaag 245
Db 241 caaag 245
RESULT 2
AAH34463
ID AAH34463 standard; cDNA; 2236 BP.
XX
XX AAH34463;
AC
XX
XX 03-SEP-2001 (first entry)
DT
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
DE
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 15; ss.

XX
XX Homo sapiens.
OS
XX
XX W0200122920-A2.
PN
XX
XX 05-APR-2001.
PD
XX
XX 28-SEP-2000; 2000MO-US26524.
PF
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX
XX WPI: 2001-225357/24.
DR P-PSDB: AAG75058.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3190-3191; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other;
SQ
Query Match 74.3%; Score 182; DB 22; Length 2236;
Best Local Similarity 100.0%; Pred. No. 3.5e-52;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 64 ctctaagtcacgggaactgcctctgctactgtgacctgccttaactcagcagttttg 123
Db 2 ctctaagtcacgggaactgcctctgctactgtgacctgccttaactcagcagttttg 123
QY 124 ttctgggaagccctgggattctgttaactactcactgtgagtgccgaaggaacaga 183
Db 62 ttctgggaagccctgggattctgttaactactcactgtgagtgccgaaggaacaga 121
QY 184 tgaagaacatgacctcaagaagctcctgtcaatgaagaagcagctgagcgtggcaa 243
Db 122 tgaagaacatgacctcaagaagctcctgtcaatgaagaagcagctgagcgtggcaa 181
QY 244 ag 245
Db 182 ag 183
RESULT 3
AAC99109
ID AAC99109 standard; cDNA; 2229 BP.
XX
XX AAC99109;
AC

XX 09-MAR-2001 (first entry)
DT Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytosolic; neuroprotective;
KW neurotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neutral; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN MO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-579444/54.
XX
DR P-PSDB: AAB54344.
XX
PS Claim 1; Page 759-760; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytosolic,
CC neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;

Query Match 71.8%; Score 176; DB 21; Length 2229;
Best Local Similarity 100.0%; Pred. No. 4.1e-50;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 gtaacggaggaactcccttgctactgtgacgccttaccagagtttggctgg 129
|||||
DB 1 gtaacggaggaactcccttgctactgtgacgccttaccagagtttggctgg 60
|||||

OY 130 gaagcccttgatctgtctaactactactactgtgagtcgtaagggaacagatgaaga 189
|||||
DB 61 gaagcccttgatctgtctaactactactactgtgagtcgtaagggaacagatgaaga 120
|||||

OY 190 acatgacctcaaggagcttcctgtcactatgagaagaccagctgcagcctggcaag 245
|||||
DB 121 acatgacctcaaggagcttcctgtcactatgagaagaccagctgcagcctggcaag 176
|||||

RESULT 4
AAC98742
ID AAC98742 standard; cDNA; 210 BP.
XX
AC AAC98742;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:752.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytosolic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN MO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587534/55.
XX
DR P-PSDB: AAB53985.
XX
PS Claim 1; Page 1310; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytosolic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC can also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 210 BP; 54 A; 47 C; 49 G; 58 T; 2 other;

Query Match 54.3%; Score 133; DB 21; Length 210;
Best Local Similarity 98.6%; Pred. No. 9e-36; 1; Indels 1; Gaps 1;
Matches 144; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 64 ctctaagtaacggaggaactcccttgctactgtgacgccttaccagagtttgg 123
|||||

	Query Match	13.0%	Score 31.8;	DB 22;	Length 327;
	Best Local Similarity	52.7%;	Pred. No. 0.75;		
	Matches	69;	Mismatches	62;	Indels 0; Gaps 0;
QY	111 ttagcagtttctgtctctggaagccctcgatcttcgtataacctcaactagtgcgt	170			

Query Match	13.0%	Score 31.8;	DB 22;	Length 330;
Best Local Similarity	52.7%;	Pred. No. 0.75;		
Matches	69;	Conservative 0;	Mismatches 62;	Indels 0;
				Gaps 0;
QY	111	tcagcagttttgtctctcgggaagccctcgatcttcgtcatacctcattcgttagtgct	170	

Db 53 tcaccattctctctctggaagaccctccaacatigtggaataatattgtctctgtacc 112
QY 171 gaagggaacagatgaagaacatgaccccaagagcttctctcatgtggaagaccacgc 230
Db 113 aacagacaccagcaagcccccgaactcatgtattatgagtcaglaagcgccctcag 172
QY 231 tgacgcctggc 241
Db 173 gggctccctgac 183

RESULT 7

AAH48118
ID AAH48118 standard; cDNA; 2966 BP.

AC AAH48118;

DT 19-SEP-2001 (first entry)

DE Dihydroorotase 9 coding sequence.

XX Dihydroorotase 9; virucide; haemostatic; immunomodulatory;

KW antiinflammatory; cytostatic; gene therapy; malignant neoplasm;

KM haemopathy; HIV infection; immunological disease; inflammation; ss.

OS Unidentified.

PN MO200148159-A1.

PD 05-JUL-2001.

PF 25-DEC-2000; 2000WO-CN00702.

PR 27-DEC-1999; 99CN-0125387.

PA (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.

PI Mao Y, Xie Y;

DR WPI: 2001-418257/44.

DR P-PSDB; AAG64239.

PT Dihydroorotase 9 and encoded polynucleotide, applicable in diagnosis

PT and treatment of malignant neoplasm, hemopathy, HIV infection,

CC immunological diseases and various inflammation

XX Claim 6; Page 27-28; 33pp; Chinese.

CC The present sequence is the coding sequence of dihydroorotase 9.

CC Dihydroorotase 9 and its coding sequence are useful for the diagnosis and

CC treatment of malignant neoplasm, haemopathy, HIV infection, immunological

CC diseases and various inflammations.

SO Sequence 2966 BP; 567 A; 786 C; 846 G; 767 T; 0 other;

Query Match 12.8%; Score 31.4; DB 22; Length 2966;

Best Local Similarity 51.0%; Pred. No. 2.7; Mismatches 71; Indels 0; Gaps 0;

Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 11 ttctctattatctatctcctcgcattctctctgtagcagagcctctctcttaag 70
Db 658 tgtctcgtattctctcttctgctgtagtgagcgctgtagcactgctctctgtagt 717
QY 71 tcaagggaactgcctctgactctgtgacctgccttactcagcagttttgtctggg 130
Db 718 caccaggaacctagatcactctgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 777
QY 131 aagcctggtattctgctataacct 155
Db 778 gaggcctgtattcagctagagcct 802

RESULT 8
AAK61101/c

ID AAK61101 standard; cDNA; 630 BP.

AC AAK61101;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6161.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KM cytostatic; gene therapy; vaccine; metastasis; ss.

OS Homo sapiens.

PN WO200157182-A2.

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAR-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0205467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226881.

PR 22-AUG-2000; 2000US-0226886.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0228287.

PR 01-SEP-2000; 2000US-0228343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2001US-0254097.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI: 2001-483426/52.	
XX	P-P5DB; AAM88320.	
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and	
PT	metastasis -	
PS	Claim 1: SEQ ID NO 6161, 3071pp + Sequence Listing; English.	
XX		
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)	
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins, and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I), by inserting	
CC	the nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/hematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK67694 represent human immune/hematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169	
CC	represent sequences used in the exemplification of the present invention.	
XX		
SQ	Sequence 630 BP; 187 A; 134 C; 169 G; 136 T; 4 other:	
Query Match	12.7%;	Score 31; DB 22; Length 630;
Best Local Similarity	50.0%;	Pred. No. 1.9; Mismatches 76; Indels 0; Gaps 0;
Matches	76; Conservative	0; Mismatches 76; Indels 0; Gaps 0;
Y	34 gcaatcattctctgtgtagtcagccctctctctctaagtcaacgggaactgccttgtaact 93	
Db	204 GCTTTATATTCTTAAACCATATATATTCCTCCAATGCACTGTAAAGTGCTCGATTCT 145	
Y	94 tttgtagcccttactcaacgagtltttgttcytggaaagcccttgaggattcgtaataac 153	
Db	144 TGCCACCACACTTTCAGAATCAAGCTTTTAAAATTCACAAGGCAGMGTGCTGTAT 85	
Y	154 ctatcacgttagtgctcgaagggaacagatg 185	
Db	84 GTCACCTTCCTCCTGTGGCAACGGGACACACATG 53	
RESULT	9	
AAH51601/C		
ID	AAH51601 standard; DNA; 319608 BP.	
XX	AAH51601;	
XX		
DT	29-AUG-2001 (first entry)	
XX		
DE	Human chromosome 13q31-q33 genomic nucleotide sequence.	
XX		

PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2001US-0254097.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	Rosen CA, Barash SC, Ruben SM;	
PI	WPI: 2001-483426/52.	
DR	P-P5DB; AAM88320.	
XX	Nucleic acids encoding human immune/haematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and	
PT	metastasis -	
PS	Claim 1: SEQ ID NO 6161, 3071pp + Sequence Listing; English.	
XX	AAR54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins, and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I), by inserting	
CC	the nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/haematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK67694 represent human immune/haematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169	
CC	represent sequences used in the exemplification of the present invention.	
XX		
SQ	Sequence 630 BP; 187 A; 134 C; 169 G; 136 T; 4 other:	
Query Match	12.7%;	Score 31; DB 22; Length 630;
Best Local Similarity	50.0%;	Pred. No. 1.9; Mismatches 76; Indels 0; Gaps 0;
Matches	76; Conservative	0; Mismatches 76; Indels 0; Gaps 0;
Dy	34 gcaatcattctctgtgagtcagccctctctctctaagtcaacgggaactgccttgtaact 93	
Dd	204 gcttttatattctttaaaccatcatatattcttccaataatgcactgtaagtgctcgattct 145	
Dy	94 tgttagtcgcccttaactcaacgaglttttgttcctgysgaagcccttgaggattcgtcataaac 153	
Dd	144 tcscacccacaccttccagaatcaacagctttttaaaattcccacaagaacgmcttcgtctaat 85	
Dy	154 ctatcacgttagtgctctgaagggaacagatg 185	
Dd	84 gtccaccttccccgtgtggcaacgggacacacacatg 53	
RESULT	9	
AAH51601/C		
ID	AAH51601 standard; DNA; 319608 BP.	
XX	AAH51601;	
XX		
DT	29-AUG-2001 (first entry)	
DE	Human chromosome 13q31-q33 genomic nucleotide sequence.	
XX		

KW	sbgl1: g34665; sbg2: g35017; g35018; chromosome 13q31-q33; haplotype;
XV	biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
XX	
OS	Homo sapiens.
PN	MO200058510-AZ.
XX	
PD	05-OCT--2000.
XX	
PF	30-MAR-2000; 200OWO-IB00435.
XX	
PR	30-MAR-1999; 99US-0126903.
XX	
PR	30-APR-1999; 99US-0131971.
XX	
PR	30-APR-1999; 99US-0132065.
XX	
PR	14-JUL-1999; 99US-0143928.
XX	
PR	27-JUL-1999; 99US-0145915.
XX	
PR	29-JUL-1999; 99US-0146452.
XX	
PR	29-JUL-1999; 99US-0146453.
XX	
PR	28-OCT-1999; 99US-0162288.
XX	
PA	(GEST) GENSET.
PI	
PI	Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;
PI	Essieux L;
XX	
DR	WPI: 2000-619082/59.
XX	
PT	Polynucleotides comprising sequences from sbg1 and g35018 biallelic
PT	markers are used for genotyping and detecting schizophrenia or bipolar
PT	disorder and predisposition to these disorders -
XX	
PS	Claim 1; Page 409-493; 737pp; English.
XX	
CC	AHH51601 represents a human genomic nucleotide sequence comprising sbg1,
CC	g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
CC	human chromosome 13q31-q33 locus. The nucleotide sequences contain
CC	biallelic markers and polymorphisms. Sequences AHH51602 - AHH51626 and
CC	AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein
CC	products. AHH51627 - AHH51631 and AAB62916 - AAB62918 represent g35018
CC	cDNA sequences and protein products. Primers AHH51632 - AHH51699 are used
CC	to isolate sbg1 cDNAs, while sbg1 exons from different primates are
CC	represented by sequences AHH51642 - AHH51699. Nucleotide sequences of
CC	amplicons which comprise biallelic markers located on the chromosome
CC	13q31-q33 locus are represented in AHH51700 -AHH51817. Biallelic markers
CC	are represented in the sequences by degenerate/undefined base codes. PCR
CC	primers AHH51818 and AHH51819 are used in the isolation of sequences of
CC	the invention. The biallelic marker containing nucleotide sequences are
CC	used to determine the identity of the nucleotide at a biallelic marker in
CC	a sample DNA sequence. The nucleotide sequences may be labelled and used
CC	for genotyping by determining the identity of a nucleotide at a Region
CC	D-related biallelic marker in a biological sample from single or multiple
CC	subjects. By determining the frequency of a biallelic marker in a
CC	population an association between a genotype and a trait, a haplotype and
CC	a trait and a phenotype and a trait can be detected. The sequences can
CC	be used to determine a predisposition to or early onset of schizophrenia or
CC	bipolar disorder or a beneficial response to or side effects related to
CC	treatment against schizophrenia or bipolar disorder.
XX	
SQ	Sequence 319608 BP; 101600 A; 56677 C; 58335 G; 102722 T; 274 other;
	Query Match 12.6%; Score 30.8; DB 21; Length 319608;
	Best Local Similarity 54.4%; Pred. No.36;
	Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0
OY	11 ttcttcattatcctacccctgcgaactctctcgagtcagagccttctctctaag 70
DB	66134 ttttcAGATGTTTCTCCTTTATTCTTTTATTGTAAACCTCATGGTATATGTCGAG 66075
OY	71 tcaggsgaacctcccttgactgtgacctgcccttactcaagaagttttt 124
DB	66074 tctctTAATCTAACCCCTCTACTGTGAATTGCTTTCTATAGTGCGTGT 66021

result_10	AAAS09301/C	
ID	AAAS09301 standard; DNA; 319608 BP.	
XX		
AC	AAAS09301;	
XX		
DT	26-SEP-2001 (first entry)	
XX		
XX	Human schizophrenia associated gene q35030 and biallelic markers	
XX		
DE	Human; q35030; biallelic marker; A1-A71; chromosome 13q31-q33;	
XX		
KW	schizophrenia; bipolar disorder; ds.	
KW		
XX		
OS	Homo sapiens.	
XX		
XX		
Key	Location/Qualifiers	
FT	7938..7958	
FT	primer_bind	
FT	/*tag= a	
FT	/note= "Binds primer 99-27943.rp"	
FT	8297..8315	
FT	primer_bind	
FT	/*tag= b	
FT	/note= "Binds primer 99-27943-150.mis"	
FT	8304..8328	
FT	misc_binding	
FT	/*tag= c	
FT	/bound_moiety= Probe_99-27943-150	
FT	8316	
FT	misc.feature	
FT	/*tag= d	
FT	/note= "Biallelic marker A1"	
FT	complement (8317..8335)	
FT	primer_bind	
FT	/*tag= e	
FT	/note= "Binds primer 99-27943-150.mis complement	
FT	complement (8446..8465)	
FT	primer_bind	
FT	/*tag= f	
FT	/note= "Binds primer 99-27943.pu complement"	
FT	21365..21385	
FT	/*tag= g	
FT	/note= "Binds primer 99-27935.rp"	
FT	21653..21671	
FT	primer_bind	
FT	/*tag= h	
FT	/note= "Binds primer 99-27935-193.mis"	
FT	21660..21684	
FT	misc_binding	
FT	/*tag= i	
FT	/bound_moiety= Probe_99-27935-193	
FT	21672	
FT	misc.feature	
FT	/*tag= j	
FT	/note= "Biallelic marker A2"	
FT	complement (21673..21691)	
FT	primer_bind	
FT	/*tag= k	
FT	/note= "Binds primer 99-27935-193.mis complement	
FT	complement (21845..21864)	
FT	primer_bind	
FT	/*tag= l	
FT	/note= "Binds primer 99-27935.pu complement"	
FT	65463..65471	
FT	primer_bind	
FT	/*tag= m	
FT	/note= "Binds primer 8-128.pu"	
FT	65466..65484	
FT	primer_bind	
FT	/*tag= n	
FT	/note= "Binds primer 8-128-33.mis"	
FT	65473..65497	
FT	misc_binding	
FT	/*tag= o	
FT	/bound_moiety= Probe_8-128-33	
FT	65485	
FT	misc.feature	
FT	/*tag= p	
FT	/note= "Biallelic marker A3"	
FT	complement (65486..65504)	
FT	primer_bind	
FT	/*tag= q	
FT	/note= "Binds primer 8-128-33.mis complement"	
FT	complement (65856..65874)	
FT	primer_bind	
FT	/*tag= r	
FT	/note= "Binds primer 8-128.rp complement"	
FT	95034..95053	
FT	primer_bind	
FT	/*tag= s	

```
FT primer_bind /note="binds primer 99-31960.pu"
FT 95377..95395
FT /tag- t
FT /note="binds primer 99-31960-363.mis"
FT 95384..95408
FT /tag- u
FT /bound_moiety= Probe_99-31960-363
FT 95396
FT /tag- v
FT /note="Biallelic marker A4"
FT complement (95397..95415)
FT /tag- w
FT /note="binds primer 99-31960-363.mis complement"
FT complement (95543..95563)
FT /tag- x
FT /note="binds primer 99-31960.rp complement"
FT 107022..107040
FT /tag- y
FT /note="binds primer 99-24656.pu"
FT 107262..107280
FT /tag- z
FT /note="binds primer 99-24656-260.mis"
FT 107269..107293
FT /tag- aa
FT /bound_moiety= Probe_99-24656-260
FT 107281
FT /tag- ab
FT /note="Biallelic marker A5"
FT complement (107282..107300)
FT /tag- ac
FT /note="binds primer 99-24656-260.mis complement"
FT complement (107495..107513)
FT /tag- ad
FT /note="binds primer 99-24656.rp complement"
FT 160279..160298
FT /tag- ae
FT /note="binds primer 99-24639.rp"
FT 160621..160639
FT /tag- ai
FT /note="binds primer 99-24639-163.mis"
FT 160628..160652
FT /tag- ag
FT /bound_moiety= Probe_99-24639-163
FT 160640
FT /tag- ah
FT /note="Biallelic marker A6"
FT complement (160641..160659)
FT /tag- aj
FT /note="binds primer 99-24639-163.mis complement"
FT 160770..160787
FT /tag- ak
FT /note="binds primer 99-24639.pu complement"
FT 160857..160875
FT /tag- al
FT /note="binds primer 99-24634-108.mis"
FT 160864..160888
FT /tag- am
FT /bound_moiety= Probe_99-24634-108
FT 160876
FT /tag- an
FT /note="Biallelic marker A7"
FT complement (160877..160895)
FT /tag- ao
FT /note="binds primer 99-24634-108.mis complement"
FT complement (161240..161257)
FT /tag- ap
FT /note="binds primer 99-24634.rp complement"
FT 168813..168830
FT /tag- aq
FT /note="binds primer 99-7652.pu"
FT 168935..168973
FT /tag- ar
FT /note="binds primer 99-7652-162.mis"
FT 168962..168986
FT /tag- as
FT /bound_moiety= Probe_99-7652-162
FT 168974
FT /tag- at
FT /note="Biallelic marker A8"
FT complement (168975..168993)
FT /tag- au
FT /note="binds primer 99-7652-162.mis complement"
FT complement (169331..169351)
FT /tag- av
FT /note="binds primer 99-7652.rp complement"
FT 170666..170686
FT /tag- aw
FT /note="binds primer 99-16100.pu"
FT 170791..170809
FT /tag- ax
FT /note="binds primer 99-16100-147.mis"
FT 170798..170822
FT /tag- ay
FT /bound_moiety= Probe_99-16100-147
FT 170810
FT /tag- az
FT /note="Biallelic marker A9"
FT complement (170811..170829)
FT /tag- ba
FT /note="binds primer 99-16100-147.mis complement"
FT complement (171153..171173)
FT /tag- bb
FT /note="binds primer 99-16100.rp complement"
FT 173065..173085
FT /tag- bc
FT /note="binds primer 99-5862.rp"
FT 173339..173357
FT /tag- bd
FT /note="binds primer 99-5862-167.mis"
FT 173346..173370
FT /tag- be
FT /bound_moiety= Probe_99-5862-167
FT 173358
FT /tag- bf
FT /note="Biallelic marker A10"
FT complement (173359..173377)
FT /tag- bg
FT /note="binds primer 99-5862-167.mis complement"
FT complement (173495..173514)
FT /tag- bh
FT /note="binds primer 99-5862.pu complement"
FT 189753..189771
FT /tag- bi
FT /note="binds primer 99-5919.pu"
FT 189938..189956
FT /tag- bj
FT /note="binds primer 99-5919-215.mis"
FT 189938..189956

Query Match 12.6%; Score 30.8; DB 22; Length 319608;
Best Local Similarity 54.4%; Pred. No. 36;
Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 11 ttctcattatctatcctcctcgtactctctgagtcagagcctctctcctaag 70
DB 66134 ttttcaagtattgtttcccttattgtttttattgtaaacctcagtgatgctgcgag 66075
QY 71 ttcagggaactgcctctgactctgagcctcttactcagcaggttttct 124
DB 66074 tcttcaatctaccctcctatctgtaattgtttcttctatgctgctgctc 66021

RESULT 11
AAK72739/c
```

ID AAK72739 standard; DNA; 2464 BP.
XX
AC AAK72739;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27551.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytosolic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
XX WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
XX PT useful for preventing, diagnosing and/or treating cancers and
XX PT metastasis -
XX
XX
XX Disclosure; SEQ ID NO 27551; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX CC amino acid sequences given in AAM82170 to AAM91971. (I) have cytostatic
XX CC activity, and can be used in gene therapy and vaccine production. (I)
XX CC proteins and polynucleotides may be used in the prevention, diagnosis and
XX CC treatment of diseases associated with inappropriate (I) expression. For
XX CC example, they may be used to treat disorders associated with decreased
XX CC expression by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of (I) by expressing inactive proteins or to
XX CC supplement the patient's own production of (I). Additionally, (I)
XX CC polynucleotides may be used to produce the secreted (I), by inserting
XX CC the nucleic acids into a host cell and culturing the cell to express the
XX CC protein. (I) proteins and polynucleotides may be used to prevent,
XX CC diagnose and treat immune/haematopoietic-related diseases, especially
XX CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX CC to AAK87694 represent human immune/haematopoietic antigen genomic
XX CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 2464 BP; 693 A; 560 C; 662 G; 549 T; 0 other;
XX
XX
XX Query Match 12.4%; Score 30.4; DB 22; Length 2464;
XX Best Local Similarity 50.0%; Pred. No. 5.6;
XX Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
XX
XX
XX 34 gcattactctctgagtcagagcctctctcttaagtcaggaactgccttgact 93
XX Db 1184 GCTTTTATCTTTAACCCATATATCTTCCAAATGCACTGTAAGTCTGAGTTCT 1125
XX
XX 94 tctgacctgccttactcagcagcttttctctggaagccctggaactctgctaact 153
XX Db 1124 TGGCACCACACTTCAGATCAAGCTTTTAAATTCACAAAGCAAGCTCTGCTAT 1065
XX
XX 154 ctatcactgtagtgctgtaaggaacagatg 185
XX Db 1064 GTCACCTTCCCTGTGCGACAGGCGCACCATG 1033
XX
XX
XX RESULT 12
XX AAK72740/C
XX ID AAK72740 standard; DNA; 3425 BP.
XX
XX AAK72740;
XX
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27552.
XX DE Human immune/haematopoietic; immune/haematopoietic antigen; cancer;
XX XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX

OS Homo sapiens.
XX
XX
XX W0200157182-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001MO-US01354.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184654.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226688.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227189.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 01-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.
XX PR 08-SEP-2000; 2000US-0232080.
XX PR 08-SEP-2000; 2000US-0232081.
XX PR 12-SEP-2000; 2000US-0231968.
XX PR 14-SEP-2000; 2000US-0232397.
XX PR 14-SEP-2000; 2000US-0232398.
XX PR 14-SEP-2000; 2000US-0232399.
XX PR 14-SEP-2000; 2000US-0232400.
XX PR 14-SEP-2000; 2000US-0232401.
XX PR 14-SEP-2000; 2000US-0233063.
XX PR 14-SEP-2000; 2000US-0233064.
XX PR 14-SEP-2000; 2000US-0233065.
XX PR 21-SEP-2000; 2000US-0234223.
XX PR 21-SEP-2000; 2000US-0234274.
XX PR 25-SEP-2000; 2000US-0234997.
XX PR 25-SEP-2000; 2000US-0234998.
XX PR 26-SEP-2000; 2000US-0235484.

PR	17-SEP-2000	2000US-023589.4	
PR	27-SEP-2000	2000US-023589.6	
PR	29-SEP-2000	2000US-023632.7	
PR	29-SEP-2000	2000US-023636.7	
PR	29-SEP-2000	2000US-023636.8	
PR	29-SEP-2000	2000US-023636.9	
PR	29-SEP-2000	2000US-023637.0	
PR	02-OCT-2000	2000US-023680.2	
PR	02-OCT-2000	2000US-023703.7	
PR	02-OCT-2000	2000US-023703.8	
PR	02-OCT-2000	2000US-023703.9	
PR	02-OCT-2000	2000US-023704.0	
PR	20-OCT-2000	2000US-024178.6	
PR	20-OCT-2000	2000US-024178.7	
PR	20-OCT-2000	2000US-024180.8	
PR	20-OCT-2000	2000US-024180.9	
PR	20-OCT-2000	2000US-024182.6	
PR	01-NOV-2000	2000US-024464.7	
PR	01-NOV-2000	2000US-024464.7	
PR	08-NOV-2000	2000US-024647.5	
PR	08-NOV-2000	2000US-024647.6	
PR	08-NOV-2000	2000US-024647.7	
PR	08-NOV-2000	2000US-024647.8	
PR	08-NOV-2000	2000US-024652.3	
PR	08-NOV-2000	2000US-024652.4	
PR	08-NOV-2000	2000US-024652.5	
PR	08-NOV-2000	2000US-024652.6	
PR	08-NOV-2000	2000US-024652.7	
PR	08-NOV-2000	2000US-024653.28	
PR	08-NOV-2000	2000US-024653.32	
PR	08-NOV-2000	2000US-024660.9	
PR	08-NOV-2000	2000US-024661.0	
PR	08-NOV-2000	2000US-024661.1	
PR	08-NOV-2000	2000US-024661.3	
PR	17-NOV-2000	2000US-024920.7	
PR	17-NOV-2000	2000US-024920.8	
PR	17-NOV-2000	2000US-024920.9	
PR	17-NOV-2000	2000US-024921.0	
PR	17-NOV-2000	2000US-024921.1	
PR	17-NOV-2000	2000US-024921.2	
PR	17-NOV-2000	2000US-024921.3	
PR	17-NOV-2000	2000US-024921.4	
PR	17-NOV-2000	2000US-024921.5	
PR	17-NOV-2000	2000US-024921.6	
PR	17-NOV-2000	2000US-024921.8	
PR	17-NOV-2000	2000US-024921.8	
PR	17-NOV-2000	2000US-024924.4	
PR	17-NOV-2000	2000US-024924.5	
PR	17-NOV-2000	2000US-024926.4	
PR	17-NOV-2000	2000US-024926.5	
PR	17-NOV-2000	2000US-024926.5	
PR	17-NOV-2000	2000US-024926.7	
PR	05-DEC-2000	2000US-025103.0	
PR	05-DEC-2000	2000US-025119.88	
PR	05-DEC-2000	2000US-025671.9	
PR	06-DEC-2000	2000US-025147.9	
PR	08-DEC-2000	2000US-025185.6	
PR	08-DEC-2000	2000US-025186.8	
PR	08-DEC-2000	2000US-025186.9	
PR	08-DEC-2000	2000US-025186.9	
PR	08-DEC-2000	2000US-025190.0	
PR	11-DEC-2000	2000US-025409.7	
PR	05-JAN-2001	2001US-025967.8	
XX	(HUMA-) HUMAN GENOME SCI INC		

XX	Rosen CA,	Barash SC,	Ruben SM;
PI			
XX	WP1: 2001-483426/52.		
DR			
XX			
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		
PT	useful for preventing, diagnosing and/or treating cancers and		
PT	metastasis -		
XX			
PS	Disclosure; SEQ ID NO 27552; 3071pp + Sequence Listing; English.		
XX			
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)		
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic		
CC	activity, and can be used in gene therapy and vaccine production. (I)		
CC	proteins and polynucleotides may be used in the prevention, diagnosis and		
CC	treatment of diseases associated with inappropriate (I) expression. For		
CC	example, they may be used to treat disorders associated with decreased		
CC	expression by rectifying mutations or deletions in a patient's genome		
CC	that affect the activity of (I) by expressing inactive proteins or to		
CC	supplement the patients own production of (I). Additionally, (I)		
CC	polynucleotides may be used to produce the secreted (I), by inserting the		
CC	nucleic acids into a host cell and culturing the cell to express the		
CC	protein. (II) proteins and polynucleotides may be used to prevent,		
CC	diagnose and treat immune/hematopoietic-related diseases, especially		
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703		
CC	to AAK87654 represent human immune/hematopoietic antigen genomic		
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169		
CC	represent sequences used in the exemplification of the present invention.		
XX			
SQ	Sequence 3425 BP: 916 A; 794 C; 951 G; 764 T; 0 other:		
<hr/>			
Query Match	12.4%;	Score 30.4;	DB 22; Length 3425;
Best Local Similarity	50.0%;	Pred. No. 6.5;	
Matches	76; Conservative	0; Mismatches	76; Indels 0; Gaps 0;
OY	34 gcaattactctctcgtagtcagagccctctctctctaagtacgagggaatgccttgactact 93		
Db	1184 GCTTTATTCCTTTAACCCCATCATATATTCTTCCAATGACGTGAAGTGCTCGAATTCT 1125		
OY	94 tgtgaccggccctttacctaacagcatgttttgtctcttgsgaaagccctggagattctgcataac 153		
Db	1124 TGCCACCACCACTTCGCAGATCAAGCCTTTTAAAATTCCAAAGGCCAAGGCTGGCTTGTACTAT 1065		
OY	154 ctatacatgttagtgcgtgaagggaacagatg 185		
Db	1064 GTACACTTCCCTCGTGGCAACGGGCAACCAATG 1033		
<hr/>			
RESULT 13			
AC	AAK72741/C		
ID	AAK72741 standard; DNA; 3425 BP.		
XX	AAK72741;		
DE	06-NOV-2001 (first entry)		
XX			
XX	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27553.		
KW	Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;		
OS	cytostatic; gene therapy; vaccine; metastasis; ds.		
XX	Homo sapiens.		
XX	WO200157182-A2.		
PN			
XX	09-AUG-2001.		
PD			
PF	17-JAN-2001; 2001WO-US01354.		
XX			
RR	31-JAN-2000; 2000US-0179065.		
OR	04-FEB-2000; 2000US-0180628.		
OR	24-FEB-2000; 2000US-0184664.		

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI, 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
PS Disclosure; SEQ ID NO 27553; 3071pp + Sequence Listing; English.
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 01:54:30 : Search time 94.91 Seconds
(without alignments)
634.077 Million cell updates/sec

Title: US-09-874-390-1_COPY_1_245

Perfect score: 245

Sequence: 1 attactgaggttcttctatt.....caagctgagcgctgcgaag 245

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCtUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32.4	13.2	7218	1	US-08-232-463-14 Sequence 14, Appl
2	29.2	11.9	1563	4	US-09-377-557-11 Sequence 11, Appl
3	28.8	11.8	72604	4	US-09-268-992-7 Sequence 7, Appl
4	28.6	11.7	333	2	US-08-652-816A-27 Sequence 27, Appl
5	28.4	11.6	1907	3	US-08-826-611-1 Sequence 1, Appl
6	28.2	11.5	2174	1	US-07-817-917-2 Sequence 2, Appl
7	27.8	11.3	1542	1	US-08-207-904-1 Sequence 1, Appl
8	27.8	11.3	3706	1	US-08-207-904-16 Sequence 16, Appl
9	27.2	11.1	347	4	US-08-952-061-3 Sequence 3, Appl
10	27.2	11.1	1571	4	US-08-868-435-32 Sequence 32, Appl
11	27.2	11.1	1571	4	US-08-744-231-32 Sequence 32, Appl
12	27	11.0	587	4	US-09-091-550A-2 Sequence 2, Appl
13	27	11.0	1280	4	US-09-276-531-52 Sequence 52, Appl
14	27	11.0	1441	1	US-08-136-277-18 Sequence 18, Appl
15	27	11.0	1441	3	US-08-479-403-18 Sequence 18, Appl
16	27	11.0	1441	3	US-08-835-734-18 Sequence 18, Appl
17	27	11.0	2750	1	US-08-136-277-1 Sequence 1, Appl
18	27	11.0	2750	2	US-08-479-403-1 Sequence 1, Appl
19	27	11.0	2750	2	US-08-835-734-1 Sequence 1, Appl
20	26.8	10.9	3066	1	US-08-142-439A-1 Sequence 1, Appl
21	26.8	10.9	3066	2	US-08-863-790-27 Sequence 1, Appl
22	26.8	10.9	4533	3	US-08-863-790-27 Sequence 27, Appl
23	26.8	10.9	4533	3	US-08-296-749-27 Sequence 27, Appl
24	26.4	10.8	158	4	US-09-019-095A-40 Sequence 40, Appl
25	26.4	10.8	160	4	US-09-019-095A-39 Sequence 39, Appl
26	26.4	10.8	918	3	US-08-937-271-9 Sequence 9, Appl
27	26.4	10.8	1650	4	US-08-868-373-7 Sequence 7, Appl

28	26.4	10.8	1873	4	US-09-019-095A-37 Sequence 37, Appl
29	26.4	10.8	2277	1	US-08-676-967-2 Sequence 2, Appl
30	26.4	10.8	2277	1	US-08-676-974-2 Sequence 2, Appl
31	26.4	10.8	2277	2	US-09-098-487-2 Sequence 2, Appl
32	26.4	10.8	2674	4	US-09-019-095A-1 Sequence 1, Appl
33	26.4	10.8	8700	2	US-08-392-625-16 Sequence 16, Appl
34	26.4	10.8	8700	2	US-08-466-961A-16 Sequence 16, Appl
35	26.4	10.8	8700	2	US-08-645-193B-18 Sequence 18, Appl
36	26	10.6	459	2	US-08-852-807-19 Sequence 19, Appl
37	26	10.6	2107	4	US-09-180-852-1 Sequence 1, Appl
38	26	10.6	2368	1	US-07-698-926A-1 Sequence 1, Appl
39	26	10.6	246240	2	US-08-724-394A-20 Sequence 20, Appl
40	26	10.6	246240	2	US-08-724-394A-21 Sequence 21, Appl
41	26	10.6	246240	2	US-08-724-394A-22 Sequence 22, Appl
42	25.8	10.5	1178	2	US-08-107-676-26 Sequence 26, Appl
43	25.8	10.5	1211	2	US-08-107-676-26 Sequence 2, Appl
44	25.8	10.5	2100	3	US-08-938-830-2 Sequence 2, Appl
45	25.8	10.5	2100	3	US-09-020-222-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F15
US-08-232-463-14

APPLICATION NUMBER: US 08/244,557
FILING DATE: 01-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: David W. Clough
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 28111/33308
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-652-816A-27

Query Match	11.7%	Score	28.6	DB	2	Length	333
Best Local Similarity	51.1%	Pred. No.	0.79				
Matches	67	Conservative	0	Mismatches	64	Indels	0
						Gaps	0

Oy	111	tcaacaaattttgtctcgggaagaccgaggatcttgtaataactacaacttgaagtgc	170
Dd	53	TCACCAATTTCTCGTCTGGAAGCACCCCAAATTGGGAATAAATTAAGTCTCCTGTACC	112
Oy	171	gaagaggaaacgatgtagaagaactagacctcaagagcttcctgtcataagaaagccaaagc	230
Dd	113	AACGACGCCCCGAGCAAAAGCCCCCAACTCATGATTTATGATGTCAGTAACGGGCGCCTCAG	172
Oy	231	tgacgcctcgagc	241
Dd	173	GGGTCCCTTGAC	183

RESULT 5
 US-08-826-611-1/c
 Sequence 1, Application US/08826611
 Patent No. 6031154
 GENERAL INFORMATION:
 APPLICANT: Bennett, Alan B.
 APPLICANT: Kanayama, Yoshio
 TITLE OF INVENTION: Fructokinase Genes and Their Use in
 TITLE OF INVENTION: Metabolic Engineering of Fruit Sweetness
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/826,611
 FILING DATE: 05-APR-1997
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Bastian, Kevin L.
 REGISTRATION NUMBER: 34,774
 REFERENCE/DOCKET NUMBER: 023070-077400US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1907 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

```

: MOLECULE TYPE: CDNA
:
: FEATURE:
:
: NAME/KEY: CDS
: LOCATION: 271..1314
:
: OTHER INFORMATION: /product="tomato Frk1"
US-08-826-611-1

```

Query Match	11.6%	Score	28.4	DB	3	Length	1907
Best Local Similarity	53.6%	Read	No. 2.1				
Matches	59	Mismatches	0	Indels	51	Gaps	0

QY	106	tttactcaagagattttgctctcgggaagcccggaatttcgtataactactactgag	165
Db	1291	TTTGTGCAACTGCATCTTGTGTGGAAGGAAGGATTCACACTCTTCTGTCACTGTCA	1232
QY	166	gtgtctgaagggaaacagatgaaagaacatgactcaagagagcttctctgtca	215
Db	1231	GGGAGAGCAACAACCTTGGCAAAAAGATAGCCCTCCCTTAACCGCTTTTCA	1182

RESULT 6
US-07-817-917-2/c
; Sequence 2, Application US/07817917
; Patent No. 5405763

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ADDRESSEE: Birch, Stewart/Kolasch & Birch
 STREET: 301 No. 5405/63th Washington Street
 CITY: Falls Church
 STATE: VIRGINIA
 COUNTRY: U.S.A.
 ZIP: 22046-3487
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/817,917
 FILING DATE: 19920109
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 001537/1991
 FILING DATE: 10-JAN-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Svensson, Leonard R.
 REGISTRATION NUMBER: 30,330
 REFERENCE/DOCKET NUMBER: 20-2941P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-241-1300
 TELEFAX: 703-241-2848
 TELEX: 248345
 INFORMATION FOR SEO ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2174 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 ORIGINAL SOURCE:
 ORGANISM: Arthrobacter globiformis
 STRAIN: SC-6-98-28 (Ferm BP-3618)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: from 1 to 1125
 IDENTIFICATION METHOD: E

US-07-817-917-2

Query Match	11.58;	Score 28.2;	DB 1;	Length 2174;
Best Local Similarity	64.68;	Pred. NO. 2.7;		
Matches 42;	Conservative 0;	Mismatches 23;	Indels 0;	Gaps 0;

Qy	235	gctctg	239
Db	2031	GCCTG	2027

```

1      RESULT      7
2      US-08-207-904-1
3      : Sequence 1, Application US/08207904
4      : Patent No. 5477002
5      : GENERAL INFORMATION:
6      : APPLICANT: Tuttle, AnnMarie
7      : APPLICANT: Crossland, Lytle D.
8      : TITLE OF INVENTION: Anthr- Specific cDNA Sequences, Genomic
9      : TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
10     : NUMBER OF SEQUENCES: 21
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: CIBA-GEIGY Corporation
13     : STREET: 7 Skyline Drive
14     : CITY: Hawthorne
15     : STATE: New York
16     : COUNTRY: USA
17     : ZIP: 10532
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Floppy disk
20     : COMPUTER: IBM PC compatible
21     : OPERATING SYSTEM: PC-DOS/MS-DOS
22     : SOFTWARE: Patentin Release #1.0, Version #1.25
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/08/207,904
25     : FILING DATE:
26     : CLASSIFICATION: 800
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER: US/07/908,242
29     : FILING DATE:
30     : ATTORNEY/AGENT INFORMATION:
31     : NAME: Lazar, Steven R.
32     : REGISTRATION NUMBER: 32,618
33     : REFERENCE/DOCKET NUMBER: CGC 1624
34     : TELECOMMUNICATION INFORMATION:
35     : TELEPHONE: (919)541-8615
36     : TELEFAX: (919)541-8689
37     : INFORMATION FOR SEQ. ID NO: 1:
38     : SEQUENCE CHARACTERISTICS:
39     : LENGTH: 1542 base pairs
40     : TYPE: nucleic acid
41     : STRANDEDNESS: single
42     : TOPOLOGY: linear
43     : MOLECULE TYPE: cDNA
44     : HYPOTHETICAL: NO
45     : ANTI-SENSE: NO
46     : ORIGINAL SOURCE:
47     : ORGANISM: Nicotiana tabacum
48     : INDIVIDUAL ISOLATE: Ant32
49     : FEATURE:
50     : NAME/KEY: CDS
51     : LOCATION: 66..1412
52     : US-08-207-904-1
53
54     Query Match      11.3%      Score 27.8; DB 1; Length 1542;
55     Best Local Similarity 59.5%; Pred. No. 3.2;
56
57     1-4-13
58

```

[illegible]

```

RESULT      8
US-08-207-904-16
; Sequence 16, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; APPLICANT: Crossland, Lyle D.
; TITLE OF INVENTION: Author-Specific cDNA Sequences, Genomic
; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana tabacum
; INDIVIDUAL ISOLATE: Ant32 genomic clone
; IMMEDIATE SOURCE:
; CLONE: PC1B950
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 1971..1975
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2076..3422
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2009
; OTHER INFORMATION: /note="Putative transcription
; OTHER INFORMATION: start site"
US-08-207-904-16

```



```

? TELEPHONE: 703-521-2297
? TELERAX: 703-685-0573
? TELEX: 248425 EMBON
? INFORMATION FOR SEQ ID NO: 18:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1441 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-136-277-18

Query Match          11.0%; Score 27; DB 1; Length 1441;
Best Local Similarity 53.3%; Pred. No. 5.9;
Matches 57; Conservative 0; Mismatches 50; Indels 0;

QY 34 gcatcattctctcgtgctcagagccctctctcctctaagtcagcgaagccttgccttact 93
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 GGATTGGTTCCTCAGCCCTCGGGCGCTGGTGTCTAGGCAACTGGAAACCTTGCGATGGCCCC 260

QY 94 tctgaccctgcctcttaccagcagcttttgtctcgtggaaagccctgg 140
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 259 TGAGACTGTCCGCTTTAGGCCCCCCGGGGGTGCTGCTGGGCCCTTGCTG 213

```

```

1. RESULT 15
2. US-08-479-403-18/c
3. : Sequence 18, Application US/08479403
4. : Patent No. 5869039
5. :
6. : GENERAL INFORMATION:
7. : APPLICANT: MANDEL, Jean-Louis
8. : APPLICANT: AUBOURG, Patrick
9. : APPLICANT: MOSSER, Jean
10. : APPLICANT: SARDE, Claude
11. : TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
12. : NUMBER OF SEQUENCES: 23
13. : CORRESPONDENCE ADDRESS:
14. : ADDRESSEE: Young & Thompson
15. : STREET: 745 South 23rd Street
16. : CITY: Arlington
17. : STATE: VA
18. : COUNTRY: USA
19. : ZIP: 22202
20. :
21. : COMPUTER READABLE FORM:
22. : MEDIUM TYPE: Floppy disk
23. : COMPUTER: IBM PC compatible
24. : OPERATING SYSTEM: PC-DOS/MS-DOS
25. : SOFTWARE: PatentIn Release #1.0, Version #1.25
26. :
27. : CURRENT APPLICATION DATA:
28. : APPLICATION NUMBER: US/08/479,403
29. : FILING DATE: 07-JUN-1995
30. : CLASSIFICATION: 514
31. :
32. : ATTORNEY/AGENT INFORMATION:
33. : NAME: PATCH, Andrew J.
34. :
35. : REGISTRATION NUMBER: 32,925
36. : REFERENCE/DOCKET NUMBER: B2272DIY
37. : TELECOMMUNICATION INFORMATION:
38. : TELEPHONE: 703-521-2297
39. : TELEFAX: 703-685-0573
40. :
41. : INFORMATION FOR SEQ ID NO: 18:
42. : SEQUENCE CHARACTERISTICS:
43. : LENGTH: 1441 base pairs
44. : TYPE: nucleic acid
45. : STRANDEDNESS: single
46. : TOPOLOGY: linear
47. :
48. : MOLECULE TYPE: DNA (genomic)
49. :
50. : US-08-479-403-18

```

Best Local Similarity	53.3%	Pred. No.	5.9
Matches	57;	Conservative	0;
		Mismatches	50;
		Indels	0;
		Gaps	0;

34 gcatlacttctctgagtcagagcctcttctctctctaaagtcacgaggaactgcctctact 93

Db 319 GGATGTGTGCTCTGACCTGGGCTGTCTTAGGCACTGGAACCTTGGATGGCCC 260

94 tgtgacctgccttactcagcagttttgttctggaagccctgg 140

Db 259 TGAGACTGTGCGCTT⁺TAGGCCCCCGGGGTGGTGCCTGGGGCCCCCTGCTG 213

Search completed: September 26, 2002, 04:22:19
Job time: 8869 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 01:51:15 : Search time 3532.1 Seconds
(without alignments)
936.201 Million cell updates/sec

Title: US-09-874-390-1_COPY_1_245

Perfect score: 245
Sequence: 1 attactgggttcctatt.....caagctgacgcctgcgaag 245

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	182	74.3	904	10	BG469649 602534110
2	179	73.1	671	10	BG750604 602708441
3	175	71.4	301	9	AA295520 EST100689
4	174	71.0	826	10	BG821370 602724814
5	172	70.2	912	10	BG469141 602511125
6	170	69.4	433	9	AA315469
7	170	69.4	871	10	BG386293 602455486
8	158	64.5	955	10	BG328280 602427256
9	157.8	64.4	503	9	AM603640 CMO-CN003
10	156.8	64.0	736	10	BE617113 601441649
11	151.8	62.0	300	9	AA345780 EST51885
12	141	57.6	297	9	AU099084 AU099084
13	138.2	56.4	850	10	BG82181 602726335
14	134	54.7	717	9	AM603669 CMO-CN003
15	131	53.5	803	10	BG469804 602534325
16	127	51.8	702	10	BG248304 602400340
17	124.4	50.8	809	10	BE872889 601450737

18	122	49.8	583	9	AM964294 EST376367
19	106	43.3	978	10	BG171515 602322061
20	105.4	43.0	882	10	BG171085 602324130
21	82.8	33.8	763	10	BT826675 603077437
22	80	32.7	1234	10	BG824793 602728784
23	73.4	30.0	830	10	BE868512 601444525
24	72	29.4	920	10	BG386247 602455439
25	72	29.4	1003	10	BG385575 602453872
26	61	24.9	864	10	BG821420 602724876
27	61	24.9	872	10	BG747491 602704606
28	56.8	23.2	629	10	BG167918 602340003
29	35.6	14.5	954	10	BE671130 602150958
30	35	14.3	451	12	A2480015 1M0301110
31	34.8	14.2	857	10	BG721017 602692685
32	34.6	14.1	778	12	A2092362 PRC1-23-2
33	34.6	14.1	881	12	BH163660 ENTSX67TR
34	34.4	14.0	403	9	A1484837 EST243098
35	34.4	14.0	567	9	A1486191 EST244512
36	34.4	14.0	577	9	AM648029 EST326483
37	34	13.9	1333	10	BM476682 AGENCOURT
38	33.8	13.8	1813	10	BE964358 601658059
39	33.4	13.6	456	12	A0378568 RPT1-11-1
40	33.4	13.6	495	12	A2508588 1M0351103
41	33.4	13.6	734	10	BG066626 H304C06-
42	33.4	13.6	885	12	A2201927 SP_0052_B
43	33.2	13.6	446	9	AA703887 ag79b04.r
44	33.2	13.6	536	10	BM328625 PRC1_24_E
45	33.2	13.6	543	12	A0303351 HS_3163_B

ALIGNMENTS

RESULT 1
LOCUS BG469649 904 bp mRNA linear EST 21-MAR-2001
DEFINITION 602534110F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4661691 5',
ACCESSION BG469649
KEYWORDS mRNA sequence.
VERSION BG469649.1 GI:13401924
SOURCE EST.
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE 1 (bases 1 to 904)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov

Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILN)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILN at:
<http://image.llnl.gov>
Plate: L1CM1461 row: c column: 04
High quality sequence stop: 604.
Location/Qualifiers
1..904

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4661691"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'-
adaptor: GGCACGAC(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)".

Query Match 73.1%; Score 179; DB 10; Length 671;
Best Local Similarity 99.5%; Pred. No. 2.5e-44;
Matches 190; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Query Match 74.3%; Score 182; DB 10; Length 904;
Best Local Similarity 100.0%; Pred. No. 3.3e-45;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 ctctaagtcagggaatgcctctgtctactgtgtactgccttactcagcaattttg 123
|||||
Db 2 CTCGAATCAACGGGAATGCCCTGTCTACTGTGACCTGCCCTTACTCAGCACTTTTG 61

QY 124 ttctgggaagccctgggattctgctaatacctcactgtagtgctggaaggaacaga 183
|||||
Db 62 TTCGGGAAGCCCTGGGATTTCTGTAATACCTACTGATGTCGAAAGGAACAGCA 121

QY 184 tgaagaacatgacctcaagaagcttctgtcaatgagaagcaagctgacgtgcaa 243
|||||
Db 122 TGAAGAATGATGACCTCAAGGAGCTTCTGTCATGAGAAGCAAGCTGACGCTGGCAA 181

QY 244 ag 245
||
Db 182 AG 183

RESULT 2

LOCUS BG750604 671 bp mRNA linear EST 15-MAY-2001
DEFINITION 602708441F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4844843 5',
mRNA sequence.

ACCESSION BG750604
VERSION BG750604.1 GI:14061257

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 671)

NIH-MGC http://mgc.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incey Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LICM1683 row: b column: 12

High quality sequence stop: 671.

Location/Qualifiers

1. 671

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4844843"

/clone_1lb="NIH_MGC_43"

/tissue_type="normal pigmented retinal epithelium"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pORF7; Site: 1; XhoI; Site: 2;
cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCAAGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library. |"

BASE COUNT 165 a 161 c 166 g 179 t

QY 55 gccctctctcttaagtcagggaatgcctctgtactgtgtactgccttactcag 114
|||||
Db 2 GCCCTTCTCTCTTAAGTCAGGGAATGCCCTGTCTACTGTGACCTGCCCTTACTCAG 61

QY 115 cagttttgtcttggaagccctgggattctgctaatacctcactgtagtgctgaag 174
|||||
Db 62 CAG-TTTTGTCTTGGAAGCCCTGGGATTTCTGTAATACCTACTGATGTCCTAG 120

QY 175 ggaacagatgaagaatgacctcaagaagcttctgtcaatgagaagcaagctgac 234
|||||
Db 121 GGAACAGATGAAGAATGACCTCAAGGAGCTTCTGTCATGAGAAGCAAGCTGAC 180

QY 235 gccctggcaag 245
|||||
Db 181 GCCTGGCAAG 191

RESULT 3

LOCUS AA295520 301 bp mRNA linear EST 18-APR-1997
DEFINITION EST100689 Pancreas tumor I Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION AA295520
VERSION AA295520.1 GI:1947854

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 301)

Adams,M.D., Kerlavage,A.R., Fleischman,R.D., Fuldner,R.A., Bult

,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White

,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,

Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald

,L.M., Fitzhugh,W.M., Fitchman,J.L., Geohagen,N.S., Glodok,A.,

Ghem,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,

Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,

Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,

Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,

Small,K.V., Spriggs,T.A., Uteyback,T.R., Weidman,J.F., Li,Y.,

Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,

Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,

Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,

Kunsch,C., Hungjun,J., Li,H., Weisener,P.S., Olsen,H., Raymond,L.,

Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon

,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and

Venter,J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

Other ESTs: THC168761

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (http://www.tigr.org/tdb/ngl/ngl.html)

Seq primer: M3 Reverse.

Location/Qualifiers

1. 301

/organism="Homo sapiens"

/db_xref="ATCC (lnhost):190965"

/db_xref="taxon:9606"

/clone.lib="Pancreas tumor I"
/dev_stage="adult"
/note="Organ: pancreas; Vector: pbluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 82 a 66 c 75 g 77 t 1 others
ORIGIN

Query Match 71.4%; Score 175; DB 9; Length 301;
Best Local Similarity 99.4%; Pred. No. 3.4e-43;
Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 70 gtcacgggaactgccttgctactgtgacgccttactcagaagtttctgtc 129
Db 1 gtcacgggaactgccttgctactgtgacgccttactcagaagtttctgtc 60
Oy 130 gaagccctgggattctgctataactatcactgtaggtcgtgaaggaaacagatgaaga 189
Db 61 GAAGCCCTGGGATTCTGTCAATACCTATCAGCTGAGGTCTGAAGGAACANATGAMGA 120
Oy 190 acatgaacctcaaggagcttcctgctcaatgaagaagcccaagcctggcgaagaag 245
Db 121 ACATGACCTCAAGAGACTTCTGTCAATGAGAGACCAAGCTGACGCTGGCAAG 176

RESULT 4
Bg821370 826 bp mRNA linear EST 22-MAY-2001
LOCUS 602724814F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4864634 5',
DEFINITION mRNA sequence.
ACCESSION Bg821370
VERSION Bg821370.1 GI:14168957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 826)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L12CM1728 row: k column: 03
High quality sequence stop: 823.

FEATURES
source
1..826
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4864634"
/clone.lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 217 a 197 c 205 g 207 t
ORIGIN

Query Match 71.0%; Score 174; DB 10; Length 826;

Best Local Similarity 100.0%; Pred. No. 9.2e-43;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 72 caaggaactgccttgctactgtgacgccttactcagaagtttctgtc 131
Db 2 CAAGGGAACCTGCTGTCAATACCTATCAGCTGAGGTCTGAAGGAACANATGAMGA 61
Oy 132 agccctgggattctgctataactatcactgtaggtcgtgaaggaaacagatgaagaac 191
Db 62 AGCCCTGGGATTCTGTCAATACCTATCAGCTGAGGTCTGAAGGAACANATGAMGA 121
Oy 192 atgaacctcaaggagcttcctgctcaatgaagaagcccaagcctggcgaagaag 245
Db 122 ATGACCTCAAGAGACTTCTGTCAATGAGAGACCAAGCTGACGCTGGCAAG 175

RESULT 5
Bg469141 912 bp mRNA linear EST 21-MAR-2001
LOCUS 602511125F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4645159 5',
DEFINITION mRNA sequence.
ACCESSION Bg469141
VERSION Bg469141.1 GI:13401416
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 912)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L12CM1418 row: b column: 08
High quality sequence stop: 705.

FEATURES
source
1..912
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4645159"
/clone.lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 240 a 214 c 231 g 227 t
ORIGIN

Query Match 70.2%; Score 172; DB 10; Length 912;
Best Local Similarity 100.0%; Pred. No. 3.9e-42;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 74 cgggaactgccttgctactgtgacgccttactcagaagtttctgtc 133
Db 2 CGGGAACCTGCTGTCAATACCTATCAGCTGAGGTCTGAAGGAACANATGAMGA 61
Oy 134 cccctggattctgctataactatcactgtaggtcgtgaaggaaacagatgaagaac 193
Db 134 cccctggattctgctataactatcactgtaggtcgtgaaggaaacagatgaagaac 193

Db 62 CCCTGGATTCGTCTATACCTATCACTGATGCTGAGGAAACAGATGAACAACAT 121

QY 194 gaacctcaaggagcttcctgtcaatgagaagcaagctgacgcctggcaag 245
|||||

Db 122 GACCTCAAGAGAGCTTCCTGTCATGAGAGACCAAGCTGACGCTGGCAAG 173

RESULT 6
AA315469
LOCUS
DEFINITION
433 bp mRNA linear EST 19-APR-1997
end, mRNA sequence.

ACCESSION
AA315469
VERSION
AA315469.1 GI:1967798
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 433)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wei,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,
L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodok,A.,
Guelm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelly,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen,H., Raymond,L.,
Weli,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,G.C.

TITLE
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence

JOURNAL
MEDLINE
Nature 377 (6547 Suppl), 3-174 (1995)

COMMENT
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 433
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):110433"
/db_xref="taxon:9606"
/clone_lib="Colon carcinoma (HCC) cell line II"
/tissue_type="colon"
/cell_type="KM12C"
/cell_line="KM12C(HCC)-parental human colon carcinoma
;Dukes B2"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 112 a 101 c 103 g 112 t 5 others
ORIGIN

Query Match 69.4%; Score 170; DB 9; Length 433;
Best Local Similarity 99.4%; Pred. No. 1.3e-41;
Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 75 gggaagacgtccctgactgtgacgtcccttaactcagcagtttctgtctgggaagc 134

Db 1 GGGAACTGCCCTTGCTACTGTGTNACCTGCCCCCTTACTACAGCTTTTGTTCGGAGC 60
|||||

QY 135 ccttgagattctgtcaatcaacctatcactgtagtgctgaaggaaacagatgaacatg 194
|||||

Db 61 CTTGGGATTCCTGCTAATACCTATCATCTAGTGTCTCTAAGGAAACAGATGAACAATG 120

QY 195 acctcaaggagcttcctgtcaatgagaagcaagctgacgcctggcaag 245
|||||

Db 121 ACCTCAAGAGAGCTTCCTGTCATGAGAGACCAAGCTGACGCTGGCAAG 171

RESULT 7
BG386293
LOCUS
DEFINITION
871 bp mRNA linear EST 12-MAR-2001
60245486F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583948 5',
mRNA sequence.

ACCESSION
BG386293
VERSION
BG386293.1 GI:13279739
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 871)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
DNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LNCM308 row: k column: 21
High quality sequence stop: 736.
Location/Qualifiers
1. 871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4583948"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(c). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 231 a 212 c 215 g 213 t
ORIGIN

Query Match 69.4%; Score 170; DB 10; Length 871;
Best Local Similarity 99.5%; Pred. No. 1.6e-41;
Matches 181; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 64 ctccaagtcacggaagcttcctgtcactgtgtacgtcccttaactcagcagttttg 123
|||||

Db 2 CTCTAAGTCACGGAGAACTGCCCTTGCTACTGTGTACTCGCCTTTACTACAGAG-TTTTG 60

QY 124 ttctgggaagccttgggattctgctaatactactactgtagtgtgtaagggaacaga 183
|||||

Db 61 TTCTGGGAAGGCCCTGGATTCCTGCTAATACCTATCACTGATGAGGAGGAACAGGA 120

QY 184 tgaagaacatgactcaagagcttctgtcactgaagaagcaagctgacgcctggcaa 243
|||||

Db	121	TGAACAATATGACCTTCAGAGACCTTCCTGTCATATGAGAAGACCAAGCTGACCCCTGGCA	180
OY	244	ag 245	
Db	181	AG 182	
RESULT	8		
LOCUS	BG328280	955 bp	mRNA
DEFINITION	602427256p1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4546813 5',	linear	EST 27-FEB-2001
ACCESSION	BG328280		
VERSION	BG328280.1	GI:13134627	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
ADHORS	1 (bases 1 to 955)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA sequencing by: NIH Intramural Sequencing Center Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LNCMI233 row: p column: 14 High quality sequence stop: 718. Location/Qualifiers		
FEATURES	source		
	1..955		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4546813"		
	/clone_lib="NIH MGC_15"		
	/tissue_type="adenocarcinoma cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"		
BASE COUNT	245 a	230 c	246 g
ORIGIN	233 t	1 others	
Query Match	64.5%;	Score 158;	DB 10; Length 955;
Best Local Similarity	98.9%;	Pred. No. 8e-38;	
Matches 180;	Conservative 0;	Mismatches 0;	Indels 2; Gaps 2.
OY	64	ctctaatgcaagggaacgcgcctctgtctactctgtgacgcgccttactcagcagttttg	123
Db	2	ctctaaagtcacgggaaacgccc-tgctactcttgtagcctgccttttactcagcag-ttttg	59
OY	124	ttctctggaagcccttgagattctgtcaataaccatactactgtagtgcctgaaggaaacaga	183
Db	60	ttctctggaagcccttgagattctgtcaataaccatactactgtagtgcctgaaggaaacaga	119
OY	184	tgaagaacatgacctcaaggagctctctgtcaatgagaagacaaagctgagccttgcaa	243
Db	120	tgaagaacatgacctcaaggagctctctgtcaatgagaagacaaagctgagccttgcaa	179
OY	244	ag 245	
Db	180	AG 181	

FEATURES	ORIGIN	BASE COUNT
LOCUS AM603640/c		
DEFINITION CM0-CN0039-260100-166-e08 CN0039 Homo sapiens CDNA, mRNA sequence.		
ACCESSION AM603640		
VERSION AM603640		
KEYWORDS EST.		
SOURCE human.		
ORGANISM Homo sapiens		
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS 1 (bases 1 to 503)		
TITLE HCGP http://www.ludwig.org.br/ORESTES .		
JOURNAL The FAPESP/LICR Human Cancer Genome Project		
COMMENT Unpublished (1999)		
Contact: Simpson A.J.G.		
Laboratory of Cancer Genetics		
Ludwig Institute for Cancer Research		
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil		
Tel: +55-11-2704922		
Fax: +55-11-2707001		
Email: asimpson@ludwig.org.br		
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL		
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM06ct2-CM0-CN0039-260100-166-e08&f3=2000-01-26&tt4=1)		
Seq primer: puc 18 forward		
High quality sequence start: 25		
High quality sequence stop: 503.		
Location/Qualifiers		
1..503		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone_lib="CN0039"		
/dev_stage="Adult"		
/note="Organ: colon,normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription ob tissue mRNA and cDNA amplification were performed under low stringency conditions."		
low stringency conditions."		
Query Match	64.4%; Score 157.8; DB 9; Length 503;	
Best Local Similarity	98.8%; Pred. No. 7.7e-38;	
Matches 159; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
OY 85 ctctctactttggccctgccccttactacagagtttcttctgggaagccctggatctc 144		
Db 503 CTTCGTACTTGTGACCTGCTCCCTTACTCAGCAGTTTGTCTGTGGAGTCTCGGATTTC 444		
OY 145 tgcataactactactactagtgtagtgctggaaggaaacagatgaagaacatgacctcaagga 204		
Db 443 TGCTAAATACATCATCTGCTGTAGTGCTGCTGAAGGAAACAGATGAAGAACATGACCTCAAGGA 384		
OY 205 gcttcctgtaataagaaagacaaagcttgaagccctgggcaag 245		
Db 383 GCTTCCTGTCAATGAGAAGACCAAGCTGACCTGGCAAG 343		
RESULT 10		
LOCUS BE617113		
DEFINITION BE617113 736 bp mRNA linear EST 20-OCT-2000		
ACCESSION BE617113		
VERSION BE617113.1 GI:9888051		

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Mammalia: Eutheria: Primates: Catarrhini: Hominiidae; Homo.
 1 (bases 1 to 736)
AUTHORS NIH-MGC <http://mgc.nhl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@email.nih.gov
 Tissue Procurement: ARCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Plate: LHAM9557 row: b column: 14
 High quality sequence start: 9
 High quality sequence stop: 670.
 Location/Qualifiers
 1..736
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3845677"
 /clone_lib="NIH-MGC-65"
 /issue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT 191 a 170 c 183 g 191 t 1 others

ORIGIN

Query Match 64.0%; Score 156.8; DB 10; Length 736;
 Best Local Similarity 98.8%; Pred. No. 1.7e-37;
 Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 86 ttgttacttgagctgaccttctactcagagctttgtcttggaagcccttgatctt 145
 |||||||
 Db 1 TTGCTACTTGAGACTGCTCCCTTTACTCAGCAGTTTGTCTGGAGACCCTGGGATTC 60

QY 146 gctaatacctacactgtaggtgctgaaggaaacagatgaagacatgacctcaaggag 205
 |||||||
 Db 61 GCTAATACCTATCTACTGTAGTCTCTGAAGGAACAGATGAAGAATGACCTCAAGGAG 120

QY 206 ctctctgcaatgagaagaccagctgacgcctgcaag 245
 |||||||
 Db 121 CTTCTCTGTCATGAGAACCAAGCTGACGCTTGCAAG 160

RESULT 11

LOCUS A0099084 300 bp mRNA linear EST 05-APR-2001

DEFINITION A0099084 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
 HEP17444 similar to Homo sapiens
 beta-16-N-acetylglucosaminyltransferase mRNA, mRNA sequence.

ACCESSION A0099084
VERSION A0099084.1 GI:13550213

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Hominiidae; Homo.
 1 (bases 1 to 300)
AUTHORS Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata
 H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo
 K., Suyama, A. and Sugano, S.
COMMENT In silico mapping of the 5'-ends of human mRNAs using full-length
 enriched and 5'-end enriched cDNA libraries constructed by

JOURNAL Oligo-capping method
COMMENT Unpublished (2001)
 Contact: Yutaka Suzuki
 Department of Virology
 Institute of Medical Science, University of Tokyo
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
 Email: ysuzuki@ims.u-tokyo.ac.jp
 Suzuki, Y., Yoshimoto-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
 S. Construction and characterization of a full length-enriched and
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
 source
 1..300
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="HEP17444"
 /clone_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 86 a 64 c 72 g 78 t

ORIGIN

Query Match 62.0%; Score 151.8; DB 9; Length 300;
 Best Local Similarity 98.7%; Pred. No. 4.6e-36;
 Matches 153; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 91 acttgacacgtgcaccttaactcagacagtttctctggaagccctggatctgctaa 150
 |||||||
 Db 1 ACTGTGACCTGCGCCCTTACTCATTTCTTCTGCGAAGCCCTGGGATTCCTGCTAA 60

QY 151 taccatcacctgtagtgctgcgaaggaaacagatgaagacatgacctcaaggagcttc 210
 |||||||
 Db 61 TACTTATCAGTGTAGTGCTGGAAGGAACAGATGAAGAATGACCTCAAGAGCTTCC 120

QY 211 tgcataatgagaagaccagctgacgcctgcaag 245
 |||||||
 Db 121 TGTCATGAGAACCAAGCTGACGCTTGCAAG 155

RESULT 12

LOCUS AA345780 297 bp mRNA linear EST 21-APR-1997

DEFINITION EST51885 gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION AA345780
VERSION AA345780.1 GI:1998058

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Hominiidae; Homo.
 1 (bases 1 to 297)

AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
 C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
 O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Val, C., Clayton, R.A.,
 Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
 L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Gloeck, A.,
 Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
 Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
 Moreno-Palancques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Sherry, R.,
 Small, K.V., Springs, T.A., Uterback, T.R., Weidman, J.F., Li, Y.,
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
 Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
 Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,
 Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
 M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
 Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl.), 3-174 (1995)
MDLINE 96026280
COMMENT Other ESTs: THC168761
 Contact: Kerlavage, AR
 Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES

SOURCE

Location/Qualifiers

1..297

/organism="Homo sapiens"

/db_xref="ATCC (Inhost):147350"

/db_xref="taxon:9606"

/clone_lib="Gall bladder I"

/sex="female"

/dev_stage="adult, 25 yrs"

/note="Organ: gall bladder; Vector: pBluescript SK-"

Site_1: EcoRI; Site_2: XhoI"

BASE COUNT

ORIGIN

84 a 59 c 76 g 75 t 3 others

Query Match 57.68; Score 141; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 9.6e-33;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 ctctaccagagtttctgtctggaagccctggatctctgctataaccctactgta 164
|||||
Db 1 ctttactcagcagttttgtctggaaagccctggatctctgctataaccctactgta 60

OY 165 ggtgctgaaggaacacagatgaagacatgacctcaagagcttctctcaatgagaaga 224
|||||
Db 61 ggtgctgaaggaacacacatgaagacatgacctcaagagcttctctcaatgagaaga 120

OY 225 ccaagctgacgcctggcgaag 245
|||||
Db 121 ccaagctgacgcctggcgaag 141

RESULT 13
Bg822181 850 bp mRNA linear EST 22-MAY-2001
LOCUS 602726335F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4865672 5',
DEFINITION mRNA sequence.

ACCESSION Bg822181
VERSION Bg822181.1 GI:14169768
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 850)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LCM171 row: f column: 09
High quality sequence stop: 696.
Location/Qualifiers

1..850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4865672"
/clone_lib="NIH_MGC_15"

/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT

ORIGIN

230 a 191 c 222 g 207 t

Query Match 56.4%; Score 138.2; DB 10; Length 850;
Best Local Similarity 94.7%; Pred. No. 9.4e-32;
Matches 143; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 95 gtgacctgcccttactacagcagtttctggaagccctggatctctgctataacc 154
|||||
Db 2 gtgacctgcccttactacagcagtttctggaagccctggatctctgctataacc 61

OY 155 tatcactgtaggtgctggaaggaacagatgaagacatgacctcaagagcttctgctc 214
|||||
Db 62 tatcactgtaggtgctggaaggaacagatgaagacatgacctcaagagcttctgctc 121

OY 215 aatgagaagaccagctgacgtcggcgaag 245
|||||
Db 122 aatgagaagaccagctgacgtcggcgaag 152

RESULT 14
AM603669 717 bp mRNA linear EST 23-MAR-2000
LOCUS CM0-CN0039-280100-168-e02 CN0039 Homo sapiens cDNA, mRNA sequence.
DEFINITION CM0-CN0039-280100-168-e02 CN0039 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM603669
VERSION AM603669.1 GI:7308410
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 717)
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0<2=CM0-CN0039-280100-168-e02<3=2000-01-28<4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 630.

FEATURES

SOURCE

Location/Qualifiers

1..717

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CN0039"

/dev_stage="Adult"

/note="Organ: colon, normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 194 a 141 c 168 g 214 t
 ORIGIN

Query Match 54.7%; Score 134; DB 9; Length 717;
 Best Local Similarity 99.3%; Pred. No. 1.8e-30;
 Matches 145; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 100 ctgaccttactcaagcagtttctgtctgggaagccctgggattctgctaatacctatca 159
 Db 14 CTGCCCTTACTCAGCAG-TTTTGTCTGGGAAGCCCTGGGATTCTGCTAATACCTATCA 72
 QY 160 ctgtaggtctgaaggaaacagatgaagacatgacctcaaggagctctctgtaatga 219
 Db 73 CTGTAGTGTCTGAAGGAACAGATGAAGAACATGACCTCAAGGAGCTTCTGCAATGA 132
 QY 220 gaagaccgaagctgacgcttgcaag 245
 Db 133 GAAGACCAAGCTGACGCTTGCAAG 158

RESULT 15

LOCUS BG469804 805 bp mRNA linear EST 21-MAR-2001
 DEFINITION 602534325F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4661670 5',
 mRNA sequence.
 ACCESSION BG469804
 VERSION BG469804.1 GI:13402079
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 805)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabps-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: L1CM1461 row: b column: 07
 High quality sequence stop: 623.
 Location/Qualifiers

FEATURES

source
 1..805
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4661670"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pORF7; Site_1: xhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGGAG(G). Site-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 205 a 205 c 191 g 204 t
 ORIGIN

Query Match 53.5%; Score 131; DB 10; Length 805;
 Best Local Similarity 94.4%; Pred. No. 1.5e-29;
 Matches 169; Conservative 0; Mismatches 5; Indels 5; Gaps 3;

QY 61 tctcttaagtcacgggaactgcctgtactctg-tgacctgaccttactcaagcagtt 119

Db 3 TCTCTTAAGTCAAGGGAACCTGCTTACTTGTGACCTCCCTTACTCAGCAGTT 62
 QY 120 ttgtcttggaagccctgggattctgctaatacctatcaagtgctgaaggaa 179
 Db 63 TATGTTCTGGGAAGCCCTGGGATTCTGCTAATACCTATCAGTGTGAAGGAAA 122
 QY 180 ca--gattgaagaacatgacct--caaggagcttctgtaatgaagaagaccagctgac 234
 Db 123 CAGGAGTGAAGAACATGACCTTCAAGAGGCTCTCTGTCAATGAGAAGACCAAGCTGAC 181

Search completed: September 26, 2002, 02:54:54
 Job time: 3819 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 02:54:54 ; Search time 3532.1 Seconds
(without alignments)
726.033 Million cell updates/sec

Title: US-09-874-390-1_COPY_246_435

Perfect score: 190
Sequence: 1 acattaaagagagcctgaa.....caggagagaagctactaaag 190

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

EST.*
1: em_estdb:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	190	100.0	583	9	AM964294 EST376367
2	190	100.0	702	10	BG248304 602400340
3	190	100.0	826	10	BG821370 602724814
4	190	100.0	864	10	BG821420 602724876
5	190	100.0	871	10	BG386293 60245486
6	188.4	99.2	850	10	BG822181 602726335
7	188.4	99.2	912	10	BG469141 602511125
8	188.4	99.2	955	10	BG328280 602427256
9	188.4	99.2	1003	10	BG385575 602453872
10	188	98.9	433	9	AA315469
11	187.4	98.6	830	10	BE868512
12	186.8	98.3	912	10	BE871910
13	178	93.7	736	10	BE617113
14	178	93.7	763	10	B1826675
15	178	93.7	872	10	BG747491
16	178	93.7	920	10	BG386247
17	178	93.7	978	10	BG171515

18	176.4	92.8	629	10	BG167918
19	176.4	92.8	671	10	BG750604
20	176.4	92.8	882	10	BG171085
c 21	169.4	89.2	337	9	AM360893
22	161.4	84.9	904	10	BG469649
23	149.8	78.8	297	9	AA345780
24	145	76.3	300	9	A0099084
25	111.6	58.7	805	10	BG469804
26	111.4	58.6	301	9	AA295520
27	102.2	53.8	766	10	B1765410
28	86.6	45.6	431	10	B1337337
29	86.6	45.6	439	10	B284104
30	84.8	44.6	582	10	B138942
31	71.6	37.7	259	10	BE748971
32	47	24.7	809	10	BE872889
33	45	23.7	875	10	BG470202
34	40	21.1	583	10	BF041857
c 35	35.2	18.5	424	10	A0196782
36	34.2	18.0	735	10	B1934209
37	33.4	17.6	1100	12	CNS00BL6
c 38	33	17.4	854	12	AG116458
39	32.6	17.2	651	12	AG045807
40	32.2	16.9	561	10	B1199475
c 41	32	16.8	611	12	AZ364585
c 42	31.6	16.6	299	10	BF545498
c 43	31.6	16.6	392	10	BG98048
44	31.4	16.5	681	12	BH261745
c 45	31.4	16.5	914	12	CNS06WYO

ALIGNMENTS

RESULT 1
LOCUS AM964294 583 bp mRNA linear EST 01-JUN-2000
DEFINITION EST376367 MAGE resequences, MAGH Homo sapiens CDNA, mRNA sequence.
ACCESSION AM964294
VERSION AM964294.1 GI:8154130
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 583)
Hegde, P., Ol, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
I, E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208

Email: johnq@tigr.org
Plate: 200
Seq primer: Reverse.

FEATURES
source Location/Qualifiers
1..583

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGH"
/note="vector: pBluescriptKm"

BASE COUNT 153 a 135 c 140 g 155 t
ORIGIN

Query Match 100.0%; Score 190; DB 9; Length 583;
Best local similarity 100.0%; Pred. No. 1.7e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atattaagaagagagcctgaactgtcttccttgagacattatgaatgctcagaataacatt 60
    |||||||
DB 123 ATATTAAAGAGAGAGCCTGGAACCTGTTCTTGACATCTTATGAATGTCAGAAAATATACCTT 182
QY 61 ttgagaggttaagaagatcaggggacatgtgtgttcacattgtgtgcacaggaacccgc 120
    |||||||
DB 183 TTGGAGGGTTAGAGATCAGGGGACATGTTGTTCACTTGTGCGCACGGAACACCGCC 242
QY 121 agcttcacttggaacagaatcacgcctgtgttaagaagatcacccttaacagagagaa 180
    |||||||
DB 243 AGCTTTCACCTGGAAACAGATCAGCGCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 302
QY 181 gctactaaag 190
    |||||||
DB 303 GCTACTAAAG 312

RESULT 2
BG248304 702 bp mRNA linear EST J3-FEB-2001
LOCUS 602400340F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546017 5',
DEFINITION mRNA sequence.
ACCESSION BG248304
VERSION BG248304.1 GI:12758119
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            Cloned into EcoRI/XhoI sites using the following 5'
            found through the I.M.A.G.E. Consortium/LNL at: image.jnl.gov
            plate: LNCM1231 row: o column: 10
            High quality sequence stop: 702.
FEATURES
    source
        1..702
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4546017"
        /clone_1ib="NIH_MGC_15"
        /tissue_type="adenocarcinoma cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site_2:
        EcoRI; cDNA made by oligo-dr priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGCACGAG(G). Size-selected >500bp for average
        insert size 1.8kb. Library constructed by Ling Hong in
        the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 186 a 163 c 175 g 178 t
ORIGIN
Query Match 100.0%; Score 190; DB 10; Length 702;
Best Local Similarity 100.0%; Pred. No. 1.8e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagagcctgaactgtcttccttgagacattatgaatgctcagaataacatt 60
    |||||||
DB 128 ATATTAAAGAGAGAGCCTGGAACCTGTTCTTGACATCTTATGAATGTCAGAAAATATACCTT 187
QY 61 ttgagaggttaagaagatcaggggacatgtgtgttcacattgtgtgcacaggaacccgc 120
    |||||||

```

```

DB 188 TTGGAGGGTTAGAGATCAGGGGACATGTTGTTCACTTGTGCGCACGGAACACCGCC 247
QY 121 agcttcacttggaacagaatcacgcctgtgttaagaagatcacccttaacagagagaa 180
    |||||||
DB 248 AGCTTTCACCTGGAAACAGATCAGCGCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 307
QY 181 gctactaaag 190
    |||||||
DB 308 GCTACTAAAG 317

RESULT 3
BG821370 826 bp mRNA linear EST 22-MAY-2001
LOCUS 602724814F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4864634 5',
DEFINITION mRNA sequence.
ACCESSION BG821370
VERSION BG821370.1 GI:14168957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 826)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            Cloned into EcoRI/XhoI sites using the following 5'
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.jnl.gov
            plate: LNCM1728 row: k column: 03
            High quality sequence stop: 823.
FEATURES
    source
        1..826
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4864634"
        /clone_1ib="NIH_MGC_15"
        /tissue_type="adenocarcinoma cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site_2:
        EcoRI; cDNA made by oligo-dr priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGCACGAG(G). Size-selected >500bp for average
        insert size 1.8kb. Library constructed by Ling Hong in
        the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 217 a 197 c 205 g 207 t
ORIGIN
Query Match 100.0%; Score 190; DB 10; Length 826;
Best Local Similarity 100.0%; Pred. No. 1.9e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagagcctgaactgtcttccttgagacattatgaatgctcagaataacatt 60
    |||||||
DB 176 ATATTAAAGAGAGAGCCTGGAACCTGTTCTTGACATCTTATGAATGTCAGAAAATATACCTT 235
QY 61 ttgagaggttaagaagatcaggggacatgtgtgttcacattgtgtgcacaggaacccgc 120
    |||||||
DB 236 TTGGAGGGTTAGAGATCAGGGGACATGTTGTTCACTTGTGCGCACGGAACACCGCC 295
QY 121 agcttcacttggaacagaatcacgcctgtgttaagaagatcacccttaacagagagaa 180
    |||||||
DB 296 AGCTTTCACCTGGAAACAGATCAGCGCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 355

```

QY 181 gctactaaag 190
|||||
Db 356 GCTACTAAG 365

RESULT 4
BG821420 864 bp mRNA linear EST 22-MAY-2001
LOCUS 602724876F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4864554 5',
DEFINITION mRNA sequence.
ACCESSION BG821420
VERSION BG821420.1 GI:14169007
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 864)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1728 row: g column: 19
High quality sequence stop: 849.

FEATURES
source location/Qualifiers
1..864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4864554"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 235 a 204 c 210 g 215 t

ORIGIN
Query Match 100.0%; Score 190; DB 10; Length 864;
Best Local Similarity 100.0%; Pred. No. 2e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaaagagagcctgaacatgtcttccttgagcacttatgaatcgaataatcctt 60
|||||
Db 62 ATATTAAAGGAGAGCCTGAACAGTCTTGGACATCTTAATGATCTGAGAAATACCTT 121

QY 61 ttggaggttagaagatcaggagacatgtgttcacattgtgcacggaacacgcgc 120
|||||
Db 122 TTGAGAGGTTAGAGATCAGGGGACATGCTGTTCACATTGTGCTGCACGAGACACGCC 181

QY 121 agctctcacttggaacacagaatcagcctgtgaaagagatcacccctaaagagagaa 180
|||||
Db 182 AGCTCTTCACTTGGAAACAGATCAGCCTTGTGAAGAGATCATCCTAAGCAGAGAGAA 241

QY 181 gctactaaag 190
|||||
Db 242 GCTACTAAG 251

RESULT 5
BG386293 871 bp mRNA linear EST 12-MAR-2001
LOCUS 602455486F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583948 5',
DEFINITION mRNA sequence.
ACCESSION BG386293
VERSION BG386293.1 GI:13279739
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1308 row: k column: 21
High quality sequence stop: 736.

FEATURES
source location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583948"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 231 a 212 c 215 g 213 t

ORIGIN
Query Match 100.0%; Score 190; DB 10; Length 871;
Best Local Similarity 100.0%; Pred. No. 2e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaaagagagcctgaacatgtcttccttgagcacttatgaatcgaataatcctt 60
|||||
Db 183 ATATTAAAGGAGAGCCTGAACAGTCTTGGACATCTTAATGATCTGAGAAATACCTT 242

QY 61 ttggaggttagaagatcaggagacatgtgttcacattgtgcacggaacacgcgc 120
|||||
Db 243 TTGAGAGGTTAGAGATCAGGGGACATGCTGTTCACATTGTGCTGCACGAGACACGCC 302

QY 121 agctctcacttggaacacagaatcagcctgtgaaagagatcacccctaaagagagaa 180
|||||
Db 303 AGCTCTTCACTTGGAAACAGATCAGCCTTGTGAAGAGATCATCCTAAGCAGAGAGAA 362

QY 181 gctactaaag 190
|||||
Db 363 GCTACTAAG 372

RESULT 6
BG822181 850 bp mRNA linear EST 22-MAY-2001
LOCUS 602726335F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4865672 5',
DEFINITION

ACCESSION mRNA sequence.
VERSION BG6822181
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1731 row: f column: 09
High quality sequence stop: 696.
Location/Qualifiers
1. 850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4865672"
/clone_lib="NIH-MGC.15"
/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 230 a 191 c 222 g 207 t
ORIGIN
Query Match 99.2%; Score 188.4; DB 10; Length 850;
Best Local Similarity 99.5%; Pred. No. 6.1e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 atataaaggagagcctgaacgttctcttgacatcttgaatgtcagaataactt 60
DB 153 ATATTAAAGAGAGACCTGAACCTGCTTCCTTGACATCTTATGAATGTCAGAAATACCTT 212
QY 61 ttgagagggttaagatcagagcagatggtgttgcacatttgcctgcacaggaacacgccc 120
DB 213 GTGAGAGGTTTGAAGATCAGGAGCATGCTTGTTCACATTGCTGCCAGAACACCGCC 272
QY 121 agtctcaacttgaaacagaatcagcctctgtgaagagatcatccctaagcaggagaa 180
DB 273 AGCTTCACCTTGAAACAGATCAGCGCTTGTGAAGAGATCATCCCTTAAGAGAGAGAA 332
QY 181 gctactaaag 190
DB 333 GCTACTAAAG 342
RESULT 7
LOCUS BG469141 912 bp mRNA linear EST 21-MAR-2001
DEFINITION 602511125F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4645159 5',
mRNA sequence.
ACCESSION BG469141
VERSION BG469141.1 GI:13401416
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1418 row: b column: 08
High quality sequence stop: 705.
Location/Qualifiers
1. 912
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4645159"
/clone_lib="NIH-MGC.15"
/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 240 a 214 c 231 g 227 t
ORIGIN
Query Match 99.2%; Score 188.4; DB 10; Length 912;
Best Local Similarity 99.5%; Pred. No. 6.2e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 atataaaggagagcctgaacgttctcttgacatcttgaatgtcagaataactt 60
DB 174 ATATTAAAGAGAGACCTGAACCTGCTTCCTTGACATCTTATGAATGTCAGAAATACCTT 233
QY 61 ttgagagggttaagatcagagcagatggtgttgcacatttgcctgcacaggaacacgccc 120
DB 234 TTGAGAGGTTTGAAGATCAGGAGCATGCTTGTTCACATTGCTGCCAGAACACCGCC 293
QY 121 agtctcaacttgaaacagaatcagcctctgtgaagagatcatccctaagcaggagaa 180
DB 294 AGCTTCACCTTGAAACAGATCAGCGCTTGTGAAGAGATCATCCCTTAAGAGAGAGAA 353
QY 181 gctactaaag 190
DB 354 GCTACTAAAG 363
RESULT 8
LOCUS BG328280 955 bp mRNA linear EST 27-FEB-2001
DEFINITION 602427256F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4546813 5',
mRNA sequence.
ACCESSION BG328280
VERSION BG328280.1 GI:13134627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC

CNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov

Plate: LNCM1233 row: P column: 14
High quality sequence stop: 718.

FEATURES

source
1. 955

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4546813"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 245 a 230 c 246 g 233 t 1 others
ORIGIN

Query Match 99.2%; Score 188.4; DB 10; Length 955;
Best Local Similarity 99.5%; Pred. No. 6.3e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atattaagagagagccttaaacctgttccttgacatcttatgaatgacgaataactt 60
DB 182 ATATTAAAGAGAGAGCGCTGAACCTGTCTTGACATCTTATGAATGTCAGAAATAACCTT 241
OY 61 ttgagagggttaagaagatacagggacatggtgttcacatttgctgcacagacacgcgc 120
DB 242 TCGGAGCGGTAGAAATGATGAGGAGCATGTTGTCACATTGCTGCCAGCAACCGCGC 301
OY 121 agcttcacttggaagacgaatcacgccttggaagagatctccctaagcagagagaa 180
DB 302 AGCTTCACTTGGAACAGAAATCACGCGCTTGGAAGAGATCATCCCTAAGCAGAGAGAA 361

OY 181 gctactaaag 190
DB 362 GCTACTAAAG 371

RESULT 9
LOCUS BG385575 1003 bp mRNA linear EST 12-MAR-2001
DEFINITION 602453872B1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582504 5',
mRNA sequence.

ACCESSION BG385575
VERSION BG385575.1 GI:13278389
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1003)
NIH-MGC http://imgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC

CNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov

Plate: LNCM1304 row: O column: 17
High quality sequence stop: 694.

FEATURES

source

1. 1003

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4582504"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 282 a 248 c 264 g 209 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 10; Length 1003;
Best Local Similarity 99.5%; Pred. No. 6.4e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atattaagagagagccttaaacctgttccttgacatcttatgaatgacgaataactt 60
DB 73 ATATTAAAGAGAGAGCGCTGAACCTGTCTTGACATCTTATGAATGTCAGAAATAACCTT 132
OY 61 ttgagagggttaagaagatacagggacatggtgttcacatttgctgcacagacacgcgc 120
DB 133 TCGGAGCGGTAGAAATGATGAGGAGCATGTTGTCACATTGCTGCCAGCAACCGCGC 192
OY 121 agcttcacttggaagacgaatcacgccttggaagagatctccctaagcagagagaa 180
DB 193 AGCTTCACTTGGAACAGAAATCACGCGCTTGGAAGAGATCATCCCTAAGCAGAGAGAA 252
OY 181 gctactaaag 190
DB 253 GCTACTAAAG 262

RESULT 10
LOCUS AA315469 433 bp mRNA linear EST 19-APR-1997
DEFINITION EST187245 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5',
end, mRNA sequence.

ACCESSION AA315469
VERSION AA315469.1 GI:1967798
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 433)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, W.-L., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodak, A., Guelm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

TITLE
JOURNAL MEDLINE
COMMENT

Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
 Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)
 96026280
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/cdb/hg1/hg1.html>)
 Seq primer: M3 Reverse.
 Location/Qualifiers

FEATURES
 source

1..433
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):110433"
 /db_xref="taxon:9606"
 /clone_lib="Colon carcinoma (HCC) cell line II"
 /issue_type="colon"
 /cell_line="KM12C"
 /note="Organ: colon; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 /note="Organ: colon; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 101 c 103 g 112 t 5 others

BASE COUNT
 112 a 101 c 103 g 112 t 5 others

ORIGIN

Query Match
 Best Local Similarity 98.9%; Score 188; DB 9; Length 433;
 Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 attttaaaggagagccgtgaacgttctcttgagacatctatgaatgcagaaataactct 60
 Db 172 ATATTAAAGAGAGCCGGAACGTGCTTGGACATCTTATGAATGCAAAAATACCTT 231
 QY 61 ttggaagggttaagaagatcaagagcatggtgttcacatttgcgtccaggaacaccgc 120
 Db 232 TTGGAGGGTTAGAAAGATCAGAGGACATGTTTTCACATTTGCTGCCACGAGACCCGCC 291
 QY 121 agcttcacttggaacagatcaagcctgtgtgaagagatcattccctaagcagagagaa 180
 Db 292 AGTTTCACTTGGNAACAGATCAGCGCTTGTGAAGAGATCAACCCCTAAGCAGAGAGAA 351
 QY 181 gctactaaag 190
 Db 352 GCTACTAAG 361

RESULT 11
BE868512
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE868512 830 bp mRNA linear EST 20-OCT-2000
 601444525F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848358 5',
 mRNA sequence.
 BE868512
 BE868512.1 GI:10317288
 EST.
 human.
 Homo sapiens

REFERENCE
AUTHORS
TITLE

NIH-MGC <http://mgc.ncl.nih.gov/>.
 1 (bases 1 to 830)
 National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM9564 row: b column: 07
 High quality sequence stop: 657.
 Location/Qualifiers

FEATURES
 source

1..830
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3848358"
 /clone_lib="NIH_MGC_65"
 /issue_type="adenoecarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 Kb. Library constructed by Life Technologies."

BASE COUNT
 213 a 207 c 210 g 210 t

ORIGIN

Query Match
 Best Local Similarity 99.5%; Score 187.4; DB 10; Length 830;
 Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 attttaaaggagagccgtgaacgttctcttgagacatctatgaatgcagaaataactct 60
 Db 262 ATATTAAAGAGAGCCGGAACGTGCTTGGACATCTTATGAATGCAAAAATACCTT 321
 QY 61 ttggaagggttaagaagatcaagagcatggtgttcacatttgcgtccaggaacaccgc 120
 Db 322 TTGGAGGGTTAGAAAGATCAGAGGACATGTTTTCACATTTGCTGCCACGAGACCCGCC 381
 QY 121 agcttcacttggaacagatcaagcctgtgtgaagagatcattccctaagcagagagaa 180
 Db 382 AGTTTCACTTGGNAACAGATCAGCGCTTGTGAAGAGATCAACCCCTAAGCAGAGAGAA 441
 QY 181 gctactaaag 189
 Db 442 GCTACTCAA 450

RESULT 12
BE871910
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE871910 912 bp mRNA linear EST 20-OCT-2000
 601447936F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3851937 5',
 mRNA sequence.
 BE871910
 BE871910.1 GI:10320686
 EST.
 human.
 Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC <http://mgc.ncl.nih.gov/>.
 1 (bases 1 to 912)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLM9573 row: g column: 10
High quality sequence stop: 624.
Location/Qualifiers

FEATURES

1..912

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3851937"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 262 a 227 c 234 g 189 t
ORIGIN

Query Match 98.3%; Score 186.8; DB 10; Length 912;
Best Local Similarity 98.9%; Pred. No. 1.9e-46;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atattaaagagagcctgaactgttccttgacatctatgaatgacgaataaccct 60
|||||
Db 40 ATATTAAAGAGAGAGCCTGAACCTGCTTCCTTGACATCTATGAAATGTCAGAAAATACCTT 99
QY 61 ttggagggttagaagatcagggagacatgtgttcacattgtctgcacaggaaccgcgc 120
|||||
Db 100 TTGGAGGGTTAGAGATCAGGGGACATGGTGTTCACATTCTCTGCCAGGAACCGGC 159
QY 121 agcttcacttggaacagacacacgcctctgtgaagagatcacccctaaagcagagagaa 180
|||||
Db 160 AGCTCTCACTTGGAACAGAAATCAGCCCTTGGAAGAGATCATCCCTAAGCAGAGAGAA 219
QY 181 gctactaaag 190
|||||
Db 220 GCTACTAAG 229

RESULT 13

BE617113 736 bp mRNA linear EST 20-OCT-2000
LOCUS 601441649P1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845677 5',
DEFINITION mRNA sequence.

ACCESSION BE617113
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@rs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLM9573 row: b column: 14
High quality sequence start: 9
High quality sequence stop: 670.

Location/Qualifiers

FEATURES

1..736

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3845677"

/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 191 a 170 c 183 g 191 t 1 others
ORIGIN

Query Match 93.7%; Score 178; DB 10; Length 736;
Best Local Similarity 99.5%; Pred. No. 8.5e-44;
Matches 189; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atattaaagagagcctgaactgttccttgacatctatgaatgacgaataaccct 60
|||||
Db 161 ATATTAAAGAGAGAGCCTGAACCTGCTTCCTTGACATCTATGAAATGTCAGAAAATACCTT 219
QY 61 ttggagggttagaagatcagggagacatgtgttcacattgtctgcacaggaaccgcgc 120
|||||
Db 220 TTGGAGGGTTAGAGATCAGGGGACATGGTGTTCACATTCTCTGCCAGGAACCGGC 279
QY 121 agcttcacttggaacagacacacgcctctgtgaagagatcacccctaaagcagagagaa 180
|||||
Db 280 AGCTCTCACTTGGAACAGAAATCAGCCCTTGGAAGAGATCATCCCTAAGCAGAGAGAA 339
QY 181 gctactaaag 190
|||||
Db 340 GCTACTAAG 349

RESULT 14

BI826675 763 bp mRNA linear EST 04-OCT-2001
LOCUS 603077437P1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5169128 5',
DEFINITION mRNA sequence.

ACCESSION BI826675
VERSION BI826675.1 GI:15938225
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 763)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@rs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLM11420 row: j column: 09
High quality sequence stop: 759.

Location/Qualifiers

FEATURES

1..763

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5169128"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb. Insert size range
0.9-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH-MGC Library."

BASE COUNT 195 a 181 c 193 g 194 t

ORIGIN

Query Match 93.7%; Score 178; DB 10; Length 763;
Best Local Similarity 99.5%; Pred. No. 8.6e-44;
Matches 189; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atattaaaggagagccctgaacatgcttccttgagacatcttatgaatgcagaaataactctt 60
|||||
Db 250 ATATTAAAGAGAGAGCCCTGAACCTGCTTCGACATCTTATGAATGTCAGAAAATACCT-T 308
|||||
QY 61 ttggaagggttagaagatcagggagacatgctgtgttcacatttgcctgcaggaacacggcc 120
|||||
Db 309 TTGGAGGGTTAGAAAGATCAGGGGACATGTTGTTTCACATTGCTGCCACGACACCGCC 368
|||||
QY 121 agcttcacttggaaacagatcacgcctctgtgaagagatcaccctaaagcagagagaa 180
|||||
Db 369 AGTCTTCACCTTGAAACAGATCAGCGCTTGTGAAGAGATCATCCCTAAGCAGAGAGAA 428
|||||
QY 181 gctactaaag 190
|||||
Db 429 GCTACTTAAG 438

RESULT 15

LOCUS BG747491 872 bp mRNA linear EST 15 MAY-2001
DEFINITION 602704606F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4858104 5',
mRNA sequence.

ACCESSION BG747491
VERSION BG747491.1 GI:14058144

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)

unpublished (1999)

Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-femail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM711 row: k column: 01

High quality sequence stop: 747.

Location/Qualifiers

1. .872

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4858104"

/clone_1b="NIH_MGC_15"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pOT87; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCAAGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT

237 a 202 c 214 g 219 t

Query Match 93.7%; Score 178; DB 10; Length 872;
Best Local Similarity 99.5%; Pred. No. 9.1e-44;
Matches 189; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atattaaaggagagccctgaacatgcttccttgagacatcttatgaatgcagaaataactctt 60
|||||
Db 62 ATATTAAAGAGAGAGCCCTGAACCTGCTTCGACATCTTATGAATGTCAGAAAATACCT-T 120
|||||
QY 61 ttggaagggttagaagatcagggagacatgctgtgttcacatttgcctgcaggaacacggcc 120
|||||
Db 121 TTGGAGGGTTAGAAAGATCAGGGGACATGTTGTTTCACATTGCTGCCACGACACCGCC 180
|||||
QY 121 agcttcacttggaaacagatcacgcctctgtgaagagatcaccctaaagcagagagaa 180
|||||
Db 181 AGTCTTCACCTTGAAACAGATCAGCGCTTGTGAAGAGATCATCCCTAAGCAGAGAGAA 240
|||||
QY 181 gctactaaag 190
|||||
Db 241 GCTACTTAAG 250

Search completed: September 26, 2002, 02:54:56
Job time: 3821 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:22:05 : Search time 4629.8 seconds
(without alignments)
858.793 Million cell updates/sec

Title: US-09-874-390-1_COPY_246_435
Perfect score: 190
Sequence: 1 atctaaagagcagctgaa.....cagagagaagctactaaag 190

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	190	100.0	2106	9	BC017032	BC017032 Homo sapi
2	190	100.0	2128	6	AR136078	AR136078 Sequence
3	190	100.0	2128	9	AF102542	AF102542 Homo sapi
4	190	100.0	187275	2	AC092755	AC092755 Homo sapi
5	188.4	99.2	2216	6	AX045249	AX045249 Sequence
6	188.4	99.2	2217	6	AX045251	AX045251 Sequence
7	188.4	99.2	2217	6	AX045253	AX045253 Sequence
8	37.4	19.7	169064	2	AC061965	AC061965 Homo sapi
9	36	18.9	150703	2	AC095344	AC095344 Rattus no
10	35.8	18.8	158098	2	AC016484	AC016484 Rattus no
11	35.6	18.7	94742	2	AC006263	AC006263 Homo sapi
12	35.2	18.5	37731	2	HSB1D7	282173 Human DNA s
13	35.2	18.5	171902	2	AC066694	AC066694 Homo sapi
14	35	18.4	178367	2	AC006115	AC006115 Homo sapi
15	34.4	18.1	164368	2	AC073986	AC073986 Homo sapi
16	34.4	18.1	169386	2	AC053485	AC053485 Homo sapi
17	34.4	18.1	181663	2	AC092415	AC092415 Homo sapi
18	34.4	18.1	195280	2	AC025081	AC025081 Homo sapi
19	34	17.9	90922	2	AC094187	AC094187 Rattus no
20	34	17.9	165007	2	AC019047	AC019047 Homo sapi
21	33.2	17.5	167676	9	AL365356	AL365356 Human DNA
22	33	17.4	317	11	HSA055W5	251719 H. sapiens (
23	33	17.4	84300	9	HSJ104A17	AL121861 Human DNA
24	33	17.4	172688	9	AL512641	AL512641 Human DNA
25	33	17.4	181570	2	AC068738	AC068738 Homo sapi
26	33	17.4	310721	2	AL627107	AL627107 Homo sapi
27	32.8	17.3	134149	2	AC013295	AC013295 Homo sapi
28	32.8	17.3	167116	9	AC009305	AC009305 Homo sapi
29	32.6	17.2	149914	9	AC073205	AC073205 Homo sapi
30	32.6	17.2	156306	2	HS873P14	AL031682 Human DNA
31	32.6	17.2	186436	2	AC068467	AC068467 Homo sapi
32	32.4	17.1	83167	9	AL590423	AL590423 Human DNA
33	32.4	17.1	110000	2	AC106860_1	Continuation (2 of
34	32.4	17.1	124000	2	AC004060	AC004060 Homo sapi
35	32.4	17.1	134354	2	AL645989	AL645989 Mus muscu
36	32.4	17.1	146851	2	AC027697	AC027697 Homo sapi
37	32.4	17.1	153936	2	AC012282	AC012282 Homo sapi
38	32.4	17.1	165606	2	AL512601	AL512601 Homo sapi
39	32.4	17.1	169638	9	AC009433	AC009433 Homo sapi
40	32.4	17.1	179000	2	AC011847	AC011847 Homo sapi
41	32.4	17.1	180915	8	GTAJ10592	AJD10592 Gullardi
42	32.4	17.1	185854	2	AC027736	AC027736 Homo sapi
43	32.4	17.1	190925	2	AL606665	AL606665 Homo sapi
44	32.4	17.1	194103	2	AC060818	AC060818 Homo sapi
45	32.4	17.1	207296	2	AL645637	AL645637 Mus muscu

ALIGNMENTS

RESULT 1
BC017032
LOCUS BC017032 2106 bp mRNA linear PRI 09-NOV-2001
DEFINITION Homo sapiens, glucosaminyl (N-acetyl) transferase 3, mucin type.
ACCESSION BC017032
VERSION BC017032.1 GI:16877561
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdex@linl.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Series: IRAC Plate: 20 Row: n Column: 9
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4758421.

FEATURES

source

1..2106
 /organism="Homo sapiens"
 /db_xref="LocusID:9245"
 /db_xref="taxon:9606"
 /clone="MGC:9086 IMAGE:3851937"
 /tissue_type="Colon, adenocarcinoma"
 /clone_id="NIH_MGC_65"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPO86"
 309..1625
 /product="glucosaminyl (N-acetyl) transferase 3, mucin
 type"
 /codon_start=1
 /protein_id="AAH17032.1"
 /db_xref="GI:16877562"
 /translation="MVQKRRLQQLHYLMAAGCYMLATVALKLSRLKCDSDHGLLES
 RESQSYCRNLIYNFLKLPARKSINSGVTRGDQDAVLQALINLEVKRRKRPFTDTH
 YLSLRDCEHFKAEKRFIOFLSKSEVEPIASVYVIAHKEIENPERLRAVYAPONTY
 CVHDEKSPETFEKAVKAIISCFPNVFTASKLVRYVYASWSRVQADLMCDLQSSV
 PKWYKLTGDEPIKSNEMVQALKMGNGRMSSEVPKPKETRMRYHEFEVBDTL
 HLTNKKDPPVNTLMTGNAYIASRPNQVLTENPKSQQLIEVYKQTPSPDEHMA
 TLQRRKRMGSPVNPMPKYDIDSMSTIARLVKMGHGDIDKAPAPSCGTHQRAICV
 YGAGDLMNMLQNHLLANKDPKVDNALQCLEEYLRKAYIGTEL"

CDS

BASE COUNT 582 a 471 c 516 g 537 t
 ORIGIN

Query Match 100.0%; Score 190; DB 9; Length 2106;
 Best Local Similarity 100.0%; Pred. No. 5,1e-49;
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 aatataaagagagcctggaactgtctcttgacatctatgaatgtcagaataacatt 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 59 ATATTAAAGAGAGCCTGAAAGCTGTTCTTGACATCTTATGAATGTCAAGAAATACCTT 118
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 61 ttggaaggttagaagaatcagaggagacatgtgttcacattgtctgcacaggaacaccgc 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 119 TTGGAGAGGTTAGAAAGATCAGGGGACATGTGTTTCACATTTGCTGCGACGGAACACCGCC 178
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 121 agcttcacttggaagaacatcagcctctgtgaagaagatcatccctaagcagagagaa 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 179 AGCTCTCAGCTTGGAACAAGATCAGCCTTGTAAGAGATCATCCTTAAGCAGAGAGAA 238
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 181 gctactaaag 190
 ||||||||||||||
 Db 239 GCTACTTAAG 248

RESULT 2
 ARI36078 2128 bp DNA linear PAT 16-JUN-2001
 LOCUS ARI36078
 DEFINITION Sequence 1 from patent US 6136580.
 ACCESSION ARI36078
 VERSION ARI36078.1 GI:14476750
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 1 Unclassified.

REFERENCE 1 (bases 1 to 2128)
 AUTHORS Fukuda, M. and Yeh, J.
 TITLE beta-1,6-N-acetylglucosaminyltransferase that forms core 2, core
 4 and I branches
 JOURNAL Patent: US 6136580-A 1 24-OCT-2000;
 FEATURES Location/Qualifiers
 source 1..2128
 /organism="unknown"

BASE COUNT 569 a 477 c 526 g 556 t
 ORIGIN

Query Match 100.0%; Score 190; DB 6; Length 2128;
 Best Local Similarity 100.0%; Pred. No. 5,1e-49;
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 aatataaagagagcctggaactgtctcttgacatctatgaatgtcagaataacatt 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 104 ATATTAAAGAGAGCCTGAAAGCTGTTCTTGACATCTTATGAATGTCAAGAAATACCTT 163
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 61 ttggaaggttagaagaatcagaggagacatgtgttcacattgtctgcacaggaacaccgc 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 164 TTGGAGAGGTTAGAAAGATCAGGGGACATGTGTTTCACATTTGCTGCGACGGAACACCGCC 223
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 121 agcttcacttggaagaacatcagcctctgtgaagaagatcatccctaagcagagagaa 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 224 AGCTCTCAGCTTGGAACAAGATCAGCCTTGTAAGAGATCATCCTTAAGCAGAGAGAA 283
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 181 gctactaaag 190
 ||||||||||||||
 Db 284 GCTACTTAAG 293

RESULT 3
 AF102542 2128 bp mRNA linear PRI 31-JAN-1999
 LOCUS Homo sapiens beta-1,6-N-acetylglucosaminyltransferase mRNA,
 complete cds.
 DEFINITION AF102542
 ACCESSION AF102542.1 GI:4204683
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2128)
 Yeh, J.C., Ong, E. and Fukuda, M.
 TITLE Molecular cloning and expression of a novel beta-1,
 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I
 branches
 JOURNAL J. Biol. Chem. 274 (5), 3215-3221 (1999)
 MEDLINE 99115671
 REFERENCE 2 (bases 1 to 2128)
 Yeh, J.C., Ong, E. and Fukuda, M.
 TITLE Direct Submission
 JOURNAL Submitted (28-OCT-1998) Glycobiology Program, The Burnham
 Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037, USA
 FEATURES Location/Qualifiers
 source 1..2128
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="brain"
 /dev_stage="fetus"
 354..1670
 /note="glycosyltransferase; core 2 Gnt-M"
 /codon_start=1
 /product="beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="AADI0824.1"
 /db_xref="GI:4204684"
 /translation="MVQKRRLQQLHYLMAAGCYMLATVALKLSRLKCDSDHGLLES
 RESQSYCRNLIYNFLKLPARKSINSGVTRGDQDAVLQALINLEVKRRKRPFTDTH
 YLSLRDCEHFKAEKRFIOFLSKSEVEPIASVYVIAHKEIENPERLRAVYAPONTY
 CVHDEKSPETFEKAVKAIISCFPNVFTASKLVRYVYASWSRVQADLMCDLQSSV

PMKYLNTCTGDEPIKSNAMVQALMKLNGRNSMESEVPPKHETPMKHYEEVROTL
HLNRKDDPPYNTLMTGTNAYIVASRDFVOHKNKPSOOLIEVMDTSPDEHMA
TLORARMPGSSVNHKPYDIDMTSTAIRVKMOCHGSDIDKGAAPACSGIHOAICV
YGAIDLMMHLLQNNHLLANKRPYVDNALOCLEBYLRYKATYGTETL

BASE COUNT 569 a 477 c 526 g 556 t

ORIGIN

Query Match 100.0%; Score 190; DB 9; Length 2128;
Best Local Similarity 100.0%; Pred. No. 5.1e-49;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atataaagagagcctgaacgtcttccttgacatctatgaatgctagaataacct 60
Db 104 ATATTAAGAGAGAGCCTGAACGTCTTCCTTGACATCTTATGAATGTCAGAAATACCTT 163
QY 61 ttggaagggttgaagatcagagacatggtgttccatcttgctgcagcagacacgcgc 120
Db 164 TTGAGAGGTTGAGAGATCAGGAGCATGTTGTCACATTTGCTGCCACGGAACACCCGC 223
QY 121 agtttcacttggaacagaaatcacgccttgtagagatcatcccttaagcagagagaa 180
Db 224 AGTCTTCACTTGGAACAGATCAGCCTGTGGAAGAGATCATCCTTAAGCAGAGAGAA 283
QY 181 gctactaaag 190
Db 284 GCTACTAAG 293

RESULT 4
AC092755 187275 bp DNA linear HTG 19-JAN-2002
LOCUS Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***
DEFINITION SEQUENCING IN PROGRESS ***, 2 ordered pieces.
AC092755 AC022480
AC092755.3 GI:18249989
VERSION HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 187275)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished

TITLE 2 (bases 1 to 187275)
JOURNAL Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.
Direct Submission
Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA

COMMENT On Jan 19, 2002 this sequence version replaced g1:15022677.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMMSC
Web site: http://chroma.mbl.washington.edu/msg_www
Contact: leecowen@systemsbiology.org

----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-Primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Insert size: -----; agarose-fp
Quality coverage: ---x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 32680: contig of 32680 bp in length
* 32681 32780: gap of unknown length
* 32781 187275: contig of 154495 bp in length.
Location/Qualifiers
1. 187275

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22"
/clone="RP11-361D15"
/clone_1b="RPC1 human BAC library 11"
/note="This clone overlaps RP11-112N19 and RP11-568G20"
BASE COUNT 53688 a 38780 c 38083 g 56624 t 100 others
ORIGIN

Query Match 100.0%; Score 190; DB 2; Length 187275;
Best Local Similarity 100.0%; Pred. No. 4.5e-49;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atataaagagagcctgaacgtcttccttgacatctatgaatgctagaataacct 60
Db 38571 ATATTAAGAGAGAGCCTGAACGTCTTCCTTGACATCTTATGAATGTCAGAAATACCTT 38630
QY 61 ttggaagggttgaagatcagagacatggtgttccatcttgctgcagcagacacgcgc 120
Db 38631 TTGAGAGGTTGAGAGATCAGGAGCATGTTGTCACATTTGCTGCCACGGAACACCCGC 38690
QY 121 agtttcacttggaacagaaatcacgccttgtagagatcatcccttaagcagagagaa 180
Db 38691 AGTCTTCACTTGGAACAGATCAGCCTGTGGAAGAGATCATCCTTAAGCAGAGAGAA 38750
QY 181 gctactaaag 190
Db 38751 GCTACTAAG 38760

RESULT 5
AX045249 2216 bp DNA linear PAT 24-NOV-2000
LOCUS Sequence 1 from Patent WO0066727.
DEFINITION AX045249
ACCESSION AX045249
VERSION AX045249.1 GI:11343799
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 2216)
Adolf, G., Heider, K.H. and Sommergruber, W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
Boehringer Ingelheim International GmbH (BE)

FEATURES
source
1. 2216
/organism="Homo sapiens"
/db_xref="taxon:9606"
5'UTR 1. 426

CDS

427..1743
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17334.1"
/db_xref="GI:11343800"
/translation="MYQMRKLCQLHYLALGCTMYLALVALKLSFRKCDSDHGLIES
RSSQYCCNLLYNFLKLPKRSINCSGVTGDOEVLQALINLEVRKRRPEFTDTH
YLSLRDEHFEKAEKRFIOEPLSKSEVEFPILVSVIHEKIEFERLLRAVAPONTY
CVHDEKSEFEKAVKAIISCFPNVFIASKLVAVYASMSRVOADLNCMEDLQSSV
PMKYFLNTGCTGDEPKISNAEVOALKMLGRMSSEVEPKHETRMKHYFEVVDTL
HLTKKDDPPVNLPMGNAYIASRDOVQVHLNPKSQQLEIEMVKDYSPDEHLMA
TLQRRAMPGSYPNHPKYDISMTSIALVYKMGQHEGDIKCAPAPPCGTHQRAICV
YGADLNMMLQNHLLANKFDPKVDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 499 c 547 g 570 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2216;
Best Local Similarity 99.5%; Pred. No. 1.6e-48;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atataaagaagagcctggaactgtcttccttgagacatctatgaatgtcagaataacctt 60
DB 177 ATATTAAAGAGGAGCCTGAAACTGTCCTTGACATCTTATGAAATGTCAGAAAATACCTT 236
OY 61 ttggaaggttaagaagatcagggagacatggttgcacattgtgcacaggaaccgcgc 120
DB 237 TTGGAGGGTTAGAAAGATCAGGGGACATGGTTGTTCAATTTGCTGCACAGAACACCGCC 296
OY 121 agcttcacttggaacagaatcacgccttctgtaagaagatcacccttaagcagaagagaa 180
DB 297 AGCTTCACTTGGAACAGATCAGCCCTTGGAAGAGATCATCCTTAAGCAGAGAGAA 356
OY 181 gctactaaag 190
DB 357 GCTACTAAAG 366

RESULT 6
AX045251
LOCUS AX045251 2217 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 3 from Patent WO0066727.
ACCESSION AX045251
VERSION AX045251.1 GI:11343801
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2217)
AUTHORS Adolf.G., Heider.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 3 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
Location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..844
1..426
427..999
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17335.1"
/db_xref="GI:11343802"
/translation="MYQMRKLCQLHYLALGCTMYLALVALKLSFRKCDSDHGLIES
RSSQYCCNLLYNFLKLPKRSINCSGVTGDOEVLQALINLEVRKRRPEFTDTH
YLSLRDEHFEKAEKRFIOEPLSKSEVEFPILVSVIHEKIEFERLLRAVAPONTY
CVHDEKSEFEKAVKAIISCFPNVFIASKLVAVYASMSRVOADLNCMEDLQSSV
PMKYFLNTGCTGDEPKISNAEVOALKMLGRMSSEVEPKHETRMKHYFEVVDTL
HLTKKDDPPVNLPMGNAYIASRDOVQVHLNPKSQQLEIEMVKDYSPDEHLMA
TLQRRAMPGSYPNHPKYDISMTSIALVYKMGQHEGDIKCAPAPPCGTHQRAICV
YGADLNMMLQNHLLANKFDPKVDNDLQCLEEYLRKAIYGTGL"
845..1744
/note="unnamed protein product"
/codon_start=1

CDS

/protein_id="CAC17336.1"
/db_xref="GI:11343803"
/translation="MRRRLTKLGCYELCMLPRTYTVSMWMMKSPETFEKAVKAIISCF
PMVFIASKLVAVYASMSRVOADLNCMEDLQSSVPMKYFLNTGCTGDEPKISNAEVO
AKMLNGRMSSEVEPKHETRMKHYFEVVDTLHLTKKDDPPVNLTMETGNAYI
VASRDOVQVHLNPKSQQLEIEMVKDYSPDEHLMA TLQRRAMPGSYPNHPKYDISDM
TSIALVYKMGQHEGDIKCAPAPPCGTHQRAICVYAGADLNMMLQNHLLANKFDPK
YDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2217;
Best Local Similarity 99.5%; Pred. No. 1.6e-48;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atataaagaagagcctggaactgtcttccttgagacatctatgaatgtcagaataacctt 60
DB 177 ATATTAAAGAGGAGCCTGAAACTGTCCTTGACATCTTATGAAATGTCAGAAAATACCTT 236
OY 61 ttggaaggttaagaagatcagggagacatggttgcacattgtgcacaggaaccgcgc 120
DB 237 TTGGAGGGTTAGAAAGATCAGGGGACATGGTTGTTCAATTTGCTGCACAGAACACCGCC 296
OY 121 agcttcacttggaacagaatcacgccttctgtaagaagatcacccttaagcagaagagaa 180
DB 297 AGCTTCACTTGGAACAGATCAGCCCTTGGAAGAGATCATCCTTAAGCAGAGAGAA 356
OY 181 gctactaaag 190
DB 357 GCTACTAAAG 366

RESULT 7
AX045253
LOCUS AX045253 2217 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 5 from Patent WO0066727.
ACCESSION AX045253
VERSION AX045253.1 GI:11343804
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2217)
AUTHORS Adolf.G., Heider.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 5 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
Location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..844
845..1744
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17337.1"
/db_xref="GI:11343805"
/translation="MRRRLTKLGCYELCMLPRTYTVSMWMMKSPETFEKAVKAIISCF
PMVFIASKLVAVYASMSRVOADLNCMEDLQSSVPMKYFLNTGCTGDEPKISNAEVO
AKMLNGRMSSEVEPKHETRMKHYFEVVDTLHLTKKDDPPVNLTMETGNAYI
VASRDOVQVHLNPKSQQLEIEMVKDYSPDEHLMA TLQRRAMPGSYPNHPKYDISDM
TSIALVYKMGQHEGDIKCAPAPPCGTHQRAICVYAGADLNMMLQNHLLANKFDPK
YDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2217;

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korva, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtenberg, O., Lieu, C., Liu, J., Liu, W.,
 Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
 Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
 Ogulu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N.,
 Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 150703)
 Worley, K.C.
 Direct Submission
 Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15625898.
 ----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GGYD
 Center clone name: CH230-11J6
 ----- Summary Statistics -----
 Assembly program: Phrap; version 0.990329first call to
 findPhrapList
 Consensus quality: 118800 bases at least Q40
 Consensus quality: 126709 bases at least Q30
 Consensus quality: 133759 bases at least Q20
 Estimated insert size: 121939; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-1p estimation
 Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draif_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 65 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 8591: contig of 8591 bp in length
 8592 gap of unknown length
 8692 contig of 5486 bp in length
 14177: gap of unknown length
 14277: gap of unknown length
 14278 contig of 3729 bp in length
 18007 gap of unknown length
 18106: gap of unknown length
 18107 contig of 3830 bp in length
 21936: gap of unknown length
 22037 contig of 5133 bp in length
 27169: gap of unknown length
 27269: gap of unknown length
 30486: contig of 3217 bp in length
 30487 gap of unknown length
 30587 contig of 3394 bp in length
 33981 gap of unknown length
 33981 gap of unknown length
 34080: gap of unknown length
 34081 contig of 3520 bp in length
 37601 gap of unknown length
 37700: contig of 3466 bp in length
 41267 gap of unknown length
 41267 contig of 3554 bp in length
 44821 gap of unknown length
 44921 contig of 3642 bp in length
 48562 gap of unknown length
 48563 contig of 2835 bp in length
 51497 gap of unknown length
 51597: gap of unknown length
 54539: contig of 2942 bp in length
 54539 gap of unknown length
 54540 contig of 2960 bp in length
 57599: gap of unknown length
 57600 contig of 3521 bp in length
 61220: gap of unknown length
 61321 contig of 2360 bp in length
 61321 gap of unknown length
 63780: contig of 2260 bp in length
 63781 gap of unknown length
 66041 gap of unknown length
 66141 contig of 2115 bp in length
 68255: gap of unknown length
 68256 gap of unknown length
 68356 contig of 3389 bp in length
 71744: gap of unknown length
 71745 contig of 2191 bp in length
 71845 gap of unknown length
 74036 contig of 1748 bp in length
 74136 gap of unknown length
 75883: gap of unknown length
 75984 contig of 2020 bp in length
 78003 gap of unknown length
 78104 contig of 2102 bp in length
 80205: gap of unknown length
 80206 gap of unknown length
 80305: contig of 2651 bp in length
 82957 gap of unknown length
 83057: gap of unknown length
 83057 contig of 2019 bp in length
 85076 gap of unknown length
 85175: gap of unknown length
 85176 contig of 1958 bp in length
 87134 gap of unknown length
 87233: gap of unknown length
 89248: contig of 2015 bp in length
 89348 gap of unknown length
 89349 contig of 2178 bp in length
 91527 gap of unknown length
 91626: gap of unknown length
 91627 contig of 2320 bp in length
 93847 gap of unknown length
 94047 contig of 1587 bp in length
 95633: gap of unknown length
 95634 gap of unknown length
 95734 contig of 1435 bp in length
 97169 gap of unknown length
 97268: gap of unknown length
 97269 contig of 1609 bp in length
 98877 gap of unknown length
 98878 contig of 2279 bp in length
 98978 gap of unknown length
 101256: gap of unknown length
 101257 contig of 1468 bp in length
 101357 gap of unknown length
 102824: gap of unknown length
 102825 contig of 1718 bp in length
 102825 gap of unknown length
 104643 contig of 1053 bp in length
 104742: gap of unknown length
 105805: contig of 1053 bp in length
 105806 gap of unknown length
 105906 contig of 1654 bp in length
 107659: gap of unknown length
 107660 gap of unknown length
 109682: contig of 2023 bp in length
 109782: gap of unknown length
 109683 contig of 1989 bp in length
 111771: gap of unknown length
 111871: gap of unknown length
 111872 contig of 1308 bp in length
 113179: gap of unknown length
 113280 contig of 1872 bp in length
 113280 gap of unknown length
 115151: gap of unknown length
 115152 contig of 1508 bp in length
 116759: gap of unknown length
 116859: gap of unknown length
 117999: contig of 1140 bp in length
 118099: gap of unknown length
 118000 contig of 1783 bp in length
 119882: contig of 1783 bp in length

```

* 119883 119982: gap of unknown length
* 119983 121000: contig of 1018 bp in length
* 121001 121100: gap of unknown length
* 121101 122530: contig of 1430 bp in length
* 122531 122630: gap of unknown length
* 122631 124087: contig of 1457 bp in length
* 124088 124187: gap of unknown length
* 124188 125929: contig of 1742 bp in length
* 125930 126030: gap of unknown length
* 126030 127298: contig of 1269 bp in length
* 127299 127398: gap of unknown length
* 127399 128963: contig of 1565 bp in length
* 128964 129063: gap of unknown length
* 129064 130295: contig of 1332 bp in length
* 130296 130395: gap of unknown length
* 130396 131716: contig of 1321 bp in length
* 131717 131816: gap of unknown length
* 131817 133191: contig of 1375 bp in length
* 133192 133291: gap of unknown length
* 133292 134948: contig of 1657 bp in length
* 134949 135048: gap of unknown length
* 135049 136777: contig of 1728 bp in length
* 136777 136877: gap of unknown length
* 136877 137885: contig of 1009 bp in length
* 137886 137985: gap of unknown length

```

Query Match 18.9% Score 36; DB 2; Length 150703;
 Best Local Similarity 60.0% Pred. No. 1;
 Matches 60; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

Oy 55 tacccttggaagggtatagaagatcaggagacatgtgttcacatttcgtcgacaggaac 114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67223 TACGTTTGGAGTGTATAGACCTTGGGGATGTGTATTTTGTGTGTCGGCCT 67164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 115 accgcagcttcactcttggaacagatacagccttgta 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67163 GGGCCAGTAGAGTCTTGAAACACTTCTCAGCTGTGTA 67124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10
 AC016484 159098 bp DNA linear HTG 14-MAR-2001
 LOCUS Homo sapiens chromosome 15 clone RP11-18H24, WORKING DRAFT
 DEFINITION
 AC016484
 AC016484.13 GI:13324773
 ACCESSION
 VERSION
 KEYWORDS
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULTOP; HTGS_ACTIVERIN.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 159098)
 REFERENCE
 AUTHORS
 Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
 Federspiel,N., Glukhov,S., Hansen,N., Herman,D.S., Hyman,R.,
 Mao,J., Komp,C., Kotler,S., Lam,B., Marathe,R., Miranda,M.,
 Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
 Southwick,A.M., Webb,C., Wilhelm,J., Yu,S. and Davis,R.W.
 Unpublished
 2 (bases 1 to 159098)
 REFERENCE
 AUTHORS
 Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
 Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
 Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D.,
 Yu,S. and Davis,R.W.
 Direct Submission
 Submitted (01-DEC-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 On Mar 14, 2001 this sequence version replaced gi:12331479.
 COMMENT
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center
 Center code: SDSTDC
 Web site: http://sequence-www.stanford.edu/group/human/

Contact: hum-info@sequence.stanford.edu
 ----- Project Information
 Center project name: 720
 Center Clone name: RP11-18H24
 ----- Summary Statistics
 Sequencing Vector: M13mp18; X02513; 99% of reads
 Sequencing Vector: plasmid; plasmid_accession: 1% of reads
 Chemistry: Dye-primer; 1% of reads
 Chemistry: Dye-terminator Big Dye; 99% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 157616 bases at least Q40
 Consensus quality: 158170 bases at least Q30
 Consensus quality: 158357 bases at least Q20
 Insert size: 156454; agarose-fp
 Insert size: 158898; sum-of-contigs
 Quality coverage: 7.9x in Q20 bases; agarose-fp
 Quality coverage: 8.2x in Q20 bases; sum-of-contigs.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1 2166: contig of 2166 bp in length
* 2167 2266: gap of unknown length
* 2267 44132: contig of 41866 bp in length
* 44133 44232: gap of unknown length
* 44233 159098: contig of 114866 bp in length.
    Location/Qualifiers
      1..159098

```

FEATURES

```

source
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="15"
  /clone="RP11-18H24"
  /clone_1lb="RPC1 human BAC library 11"
  1..2166
    misc_feature
      /note="assembly_name:Contig31"
      2267..44132
    misc_feature
      /note="assembly_name:Contig32"
      44233..159098
    misc_feature
      /note="assembly_name:Contig33"
      clone_end:SP6"
BASE COUNT 46075 a 33092 c 33649 g 46082 t 200 others
ORIGIN
```

Query Match 18.8% Score 35.8; DB 2; Length 159098;
 Best Local Similarity 55.1% Pred. No. 1.2;
 Matches 70; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

Oy 36 tcttatgaatgcagaataacaccttggaagggtagaagatcaggagacatggtgttc 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138922 TTTTGAACACCCAGAAAATAGTGTGTGAGATGTAGAGAAATGTGAAGCTTGATC 138981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 96 acattgctgcacaggaacacgcagcttcactcttggaacagatacagcctgtgtaa 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138982 ACTCTTGCTGGAACATTAATAATGGACACAGCTGCTGTGGAACATTATGACATTTCTTAA 139041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 156 gagatca 162
    |||
Db 139042 ATCATTA 139048

```

RESULT 11
 AC006263 94742 bp DNA linear PRI 01-JAN-1999
 LOCUS Homo sapiens chromosome 17, clone hC17.187_K.10, complete sequence.
 DEFINITION
 AC006263
 AC006263.1 GI:4092479
 ACCESSION
 VERSION
 KEYWORDS
 HTG.
 SOURCE
 human.

ORGANISM	Homo sapiens	repeat_region	complement(4956..5273)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	repeat_region	/rpt_family="AluSx"
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	repeat_region	complement(5316..5681)
TITLE	1 (bases 1 to 94742)	repeat_region	/rpt_family="MLT1B"
JOURNAL	Homo sapiens chromosome 17, clone hC17.187_K.10	repeat_region	5803..5889
REFERENCE	2 (bases 1 to 94742)	repeat_region	/rpt_family="L2"
AUTHORS	Unpublished	repeat_region	6260..6553
	2 (bases 1 to 94742)	repeat_region	/rpt_family="AluSg"
	Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,	repeat_region	7723..7999
	Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, D., Brown, A.,	repeat_region	/rpt_family="AluJo"
	Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,	repeat_region	8000..8044
	Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,	repeat_region	/rpt_family="CA)n"
	Donelan, L., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R.,	repeat_region	complement(8205..8506)
	Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A.,	repeat_region	/rpt_family="AluSx"
	Herena, L., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A.,	repeat_region	10280..10318
	Lehoczky, J., Macdonald, P., Marquis, N., McEwan, P., McGurk, A.,	repeat_region	/rpt_family="MLT1"
	McKernan, K., Meldrim, J., Molla, M., Morris, W., Morrow, J.,	repeat_region	10430..10478
	Mychaleckyj, J., Naylor, J., Nilloff, M., O'Connor, T., O'Donnell, P.,	repeat_region	/rpt_family="MER58A"
	Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A.,	repeat_region	complement(10479..10776)
	Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C.,	repeat_region	/rpt_family="AluSg"
	Subramanian, A., Testaye, S., Torruella-Miller, I., Vassiliev, H.,	repeat_region	10777..10950
	Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.	repeat_region	/rpt_family="MER58A"
TITLE	Submitted (30-DEC-1998) Whitehead Institute/MIT Center for Genome	repeat_region	complement(11942..12248)
JOURNAL	Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	/rpt_family="AluSx"
REFERENCE	3 (bases 1 to 94742)	repeat_region	12505..12648
AUTHORS	Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,	repeat_region	/rpt_family="L2"
	Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, D., Brown, A.,	repeat_region	/rpt_family="MADE1"
	Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,	repeat_region	complement(16984..17289)
	Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,	repeat_region	/rpt_family="AluSg"
	Donelan, L., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R.,	repeat_region	18258..18355
	Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A.,	repeat_region	/rpt_family="MIR"
	Herena, L., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A.,	repeat_region	/rpt_family="MIR"
	Lehoczky, J., Macdonald, P., Marquis, N., McEwan, P., McGurk, A.,	repeat_region	19744..19961
	McKernan, K., Meldrim, J., Molla, M., Morris, W., Morrow, J.,	repeat_region	/rpt_family="MIR"
	Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A.,	repeat_region	complement(20015..20289)
	Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C.,	repeat_region	/rpt_family="L2"
	Subramanian, A., Testaye, S., Torruella-Miller, I., Vassiliev, H.,	repeat_region	complement(22035..22411)
	Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.	repeat_region	/rpt_family="THE1C"
TITLE	Submitted (01-JAN-1999) Whitehead Institute/MIT Center for Genome	repeat_region	complement(22492..22620)
JOURNAL	Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	/rpt_family="MIR"
COMMENT	On Jan 1, 1999 this sequence version replaced g1:4079623.	repeat_region	complement(22621..22926)
	All repeats were identified using RepeatMasker: Smit, A.F.A. &	repeat_region	/rpt_family="AluSg"
	Green, P. (1996-1997)	repeat_region	24191..24327
	http://ftp.genome.washington.edu/RM/RepeatMasker.html.	repeat_region	/rpt_family="AluJo"
FEATURES	Location/Qualifiers	repeat_region	24328..24635
source	1. 94742	repeat_region	/rpt_family="AluSg"
	/organism="Homo sapiens"	repeat_region	24636..24804
	/db_xref="taxon:9606"	repeat_region	/rpt_family="AluJo"
	/chromosome="17"	repeat_region	complement(25514..25800)
	/map="17"	repeat_region	/rpt_family="AluSx"
	/clone_lib="NCIT_187_K_10"	repeat_region	25829..26097
	/complement(11..109)	repeat_region	/rpt_family="L2"
	/rpt_family="MIR"	repeat_region	26112..26211
	complement(363..637)	repeat_region	/rpt_family="L2"
	/rpt_family="AluSx"	repeat_region	28378..28417
	643..689	repeat_region	/rpt_family="CA)n"
	/rpt_family="AT_rich"	repeat_region	complement(28562..28598)
	complement(1879..2173)	repeat_region	/rpt_family="tRNA-Cys-TGC"
	/rpt_family="AluSx"	repeat_region	complement(29242..29565)
	complement(2184..2496)	repeat_region	/rpt_family="AluSc"
	/rpt_family="AluY"	repeat_region	29686..29725
	2960..3003	repeat_region	/rpt_family="AluJo"
	/rpt_family="AT_rich"	repeat_region	32426..32446
	complement(3438..3746)	repeat_region	/rpt_family="CA)n"
	/rpt_family="Aluub"	repeat_region	32450..32736
	3817..4159	repeat_region	/rpt_family="AluJo"
	/rpt_family="MER47A"	repeat_region	32737..32766
	4867..4887	repeat_region	complement(33301..33338)
	/rpt_family="AT_rich"	repeat_region	/rpt_family="A-rich"
		repeat_region	complement(34791..34833)
		repeat_region	/rpt_family="TAAA)n"
		repeat_region	36154..36312

```
repeat_region      /rpt_family="AluY"
                    36313. 36339
repeat_region      /rpt_family="(CAA)n"
                    complement(36756. 37064)
repeat_region      /rpt_family="AluSx"
                    complement(37923. 38269)
repeat_region      /rpt_family="THE1B"
                    39314. 39753
repeat_region      /rpt_family="MT1C"
                    40962. 41328
repeat_region      /rpt_family="THE1C"
                    complement(41915. 42227)
repeat_region      /rpt_family="AluSx"
                    complement(42389. 42489)
repeat_region      /rpt_family="MIR"
                    42542. 42814
repeat_region      /rpt_family="AluJb"
                    43264. 43294
repeat_region      /rpt_family="AT_rich"
                    44375. 44680
repeat_region      /rpt_family="AluJo"
                    46335. 46413
repeat_region      /rpt_family="(CA)n"
                    complement(46830. 47141)
repeat_region      /rpt_family="AluSg"
                    complement(47166. 48452)
repeat_region      /rpt_family="11MC1"
                    complement(48835. 49146)
repeat_region      /rpt_family="AluJo"
                    50833. 50862
repeat_region      /rpt_family="MIR"
                    complement(51012. 51304)
repeat_region      /rpt_family="AluJo"
                    51317. 51358
repeat_region      /rpt_family="A-rich"
                    52808. 52827
repeat_region      /rpt_family="(CAAA)n"
                    complement(53638. 53974)
repeat_region      /rpt_family="AluSg"
                    54446. 54460
repeat_region      /rpt_family="AT_rich"
                    55090. 55110
repeat_region      /rpt_family="AT_rich"
                    complement(55411. 55687)
repeat_region      /rpt_family="AluJo"
                    complement(56518. 56574)
repeat_region      /rpt_family="putrline-rich"
                    57117. 57403
```

Query Match	Similarity	18.7%	Score 35.6	DB 9	Length 94742
Best Local	Similarity	54.6%	Pred. No. 1.4		
Matches	71	Conservative	0	Mismatches	59
				Indels	0
				Gaps	0
Qy	61	tttgagaggtltaagaagatcaaggagacatggltgttcacattgtgcgcacgagacaccgc	120		
Db	32641	TCGGAGAGATTGCTTCAGCCACGACGAGATGTAAGTGTGAGTGTGACACAGAGATGACACCACTGC	32700		
Qy	121	agcttctacttggaaacagacagatcacgccttgtgaagagatcccttaagcagagagaa	180		
Db	32701	ACTCCAGCGCTGGCACAACAGACCAACCCCTATTCCAAAAAACAACCAACCAACCAAAA	32760		
Qy	181	gctactaaag 190			
Db	32761	GCAAAAAAAG 32770			
RESULT	12				
HSB1D7					
LOCUS					
DEFINITION	Human DNA sequence from clone SC22CB-1D7 on chromosome 22 contains two exons of the LARGE gene for like-glycosyltransferase (K1AA0609), ESTs, an STS and GSSs, complete sequence.				
ACCESSION	Z831173				

VERSION	282173.2	GI:6572307
KEYWORDS	HTG; glycosyltransferase; KIAA0609; LARGE.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	Whiteley, M.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquereyesanger.ac.uk Clone requests: clonerequestesanger.ac.uk	
COMMENT	On Dec 13, 1999 this sequence version replaced gi:3426102.	

During sequence assembly data is compressed from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wg: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22> SC32GB-ID7 is from the human chromosome 22-specific cosmid library (SC32CB) constructed at the Sanger Centre by Mark Koss and Cordelia Langford.

FEATURES

VECTOR: lawr1st16

This sequence is the entire insert of clone SC22CB-1D7.

Location/Qualifiers

```

source
1. .37731
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="SC22CB-1D7"
/clone_11b="SC22CB"
1. .77
/note="AluY repeat: matches 2. .78 of consensus"
526. .826
/note="AluSg repeat: matches 1. .310 of consensus"
complement(608..1161)
/note="match: GSS: Em:BI8157"
1144. .1295
/note="MIR repeat: matches 27. .189 of consensus"
1296. .1591
/note="match: STS: Em:G03780
match: STS: Em:G03780"
1305. .1664
/note="match: GSS: Em:BI8152"
1563. .1634
/note="L2 repeat: matches 2607. .2666 of consensus"
1645. .1699
/note="MIR11 repeat: matches 85. .139 of consensus"
complement(1657..2008)
/note="match: GSS: Em:A0401622"
1700. .1802
/note="AluSg/x repeat: matches 199. .302 of consensus"
2163. .2558
/note="match: GSS: Em:A0705678"
2302. .2413
/note="MIR repeat: matches 29. .153 of consensus"
2497. .2545
/note="MIR11 repeat: matches 96. .144 of consensus"

```

```

repeat_region 2687..2790 /note="MIR1H repeat: matches 428..534 of consensus"
repeat_region 2803..2886 /note="MIR repeat: matches 178..262 of consensus"
repeat_region 2961..3041 /note="L2 repeat: matches 2607..2675 of consensus"
repeat_region 3053..3105 /note="MIR1A1 repeat: matches 86..138 of consensus"
repeat_region 3106..3401 /note="AluY repeat: matches 1..295 of consensus"
repeat_region 3402..3614 /note="MIR1A1 repeat: matches 138..347 of consensus"
repeat_region 3622..3939 /note="L2 repeat: matches 2107..2438 of consensus"
repeat_region 3962..4144 /note="MER5A repeat: matches 1..189 of consensus"
repeat_region 4899..5194 /note="AluY repeat: matches 1..296 of consensus"
repeat_region 5457..5622 /note="MIR repeat: matches 37..214 of consensus"
repeat_region 5708..5940 /note="MIR repeat: matches 11..241 of consensus"
repeat_region 6008..6129 /note="MIR repeat: matches 100..214 of consensus"
repeat_region 6130..6395 /note="AluSc repeat: matches 34..301 of consensus"
repeat_region 6396..6487 /note="MIR repeat: matches 7..100 of consensus"
repeat_region 7028..7277 /note="L2 repeat: matches 2246..2497 of consensus"
misc_feature complement(7141..7684)
repeat_region 7521..7883 /note="match: GSS: Em:AQ699024"
repeat_region 7897..8003 /note="MIR1E repeat: matches 180..564 of consensus"
repeat_region 8459..8530 /note="L2 repeat: matches 2583..2704 of consensus"
repeat_region 8471..8565 /note="L2 repeat: matches 2678..2748 of consensus"
repeat_region 8868..9076 /note="MIR repeat: matches 165..262 of consensus"
repeat_region 9107..9269 /note="AluY repeat: matches 6..214 of consensus"
repeat_region 9564..9862 /note="MER5B repeat: matches 1..172 of consensus"
repeat_region 9564..9862 /note="AluSg repeat: matches 1..299 of consensus"
gene complement(join(10020..10183,31581..31736))
CDS /gene="LARGE"
/note="match: CDNAS: Em:AJ006278 Em:AB011181 Em:AJ007583
match: ESTs: Em:AI630787 Em:T09171 Em:AA916781 Em:AA976610
Em:T31101 Em:H55069
match: proteins: Tr:O60348 Tr:O95461 Tr:Q921M7"
/codon_start=1
/evidence=not_experimental
/product="cbid7.1 (like-glycosyltransferase (KIA0609))"
/protein_id="CAB62965.1"
/db_xref="GI:6572308"
/translation="VIHNSPKIKRYKNHVEFRLYLTLEYGONLIRLEFGPS
EADVNSENLQKQISELDEDLDYEFRRERFTVHRHLFLAHEYEPADSDTVTLVAQ
LSMD"
misc_feature complement(10020..10183)
/gene="LARGE"
/note="match: STS: Em:H55069"
repeat_region 10394..10701
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 10867..11178
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 11757..11875
/note="LIME3A repeat: matches 5993..6115 of consensus"
repeat_region 11890..12187
/note="AluYB8 repeat: matches 1..306 of consensus"

```

```

repeat_region 12248..12484 /note="LIP4 repeat: matches 5716..5952 of consensus"
repeat_region 12488..12684 /note="LIP4 repeat: matches 5948..6144 of consensus"
repeat_region 13025..13191 /note="MIR1H repeat: matches 25..204 of consensus"
repeat_region 13338..13630 /note="AluSg repeat: matches 1..291 of consensus"
repeat_region 14116..14381 /note="L2 repeat: matches 2298..2528 of consensus"
repeat_region 14922..15138 /note="MIR repeat: matches 8..261 of consensus"
repeat_region 15771..16053 /note="AluSx repeat: matches 1..283 of consensus"
repeat_region 16776..17100 /note="AluY repeat: matches 1..311 of consensus"
repeat_region 17226..17283 /note="L2 repeat: matches 2648..2705 of consensus"
repeat_region 17235..17413 /note="MIR repeat: matches 48..254 of consensus"
repeat_region 17885..18193 /note="AluSx repeat: matches 1..312 of consensus"
repeat_region 18197..18270 /note="37 copies 2 mer aa 67 conserved"
misc_feature complement(18347..19008)
/gene="LARGE"
/note="match: GSS: Em:AQ284466"
complement(18650..19035)
/gene="LARGE"
/note="match: GSS: Em:AQ167619"
repeat_region 18735..18772 /note="MIR repeat: matches 65..102 of consensus"
repeat_region 18832..19054 /note="MIR repeat: matches 15..240 of consensus"
misc_feature 19051..19425 /note="match: GSS: Em:AQ238543"
repeat_region 19218..19285 /note="MIR1J repeat: matches 103..172 of consensus"
repeat_region 19408..19480 /note="MIR1J repeat: matches 327..403 of consensus"
repeat_region 19529..19535 /note="could be 8 bp"
repeat_region 20201..20271 /note="L2 repeat: matches 2642..2710 of consensus"
repeat_region 20364..20600 /note="MIR repeat: matches 13..255 of consensus"
repeat_region 20651..20655 /note="could be 4T's"
repeat_region 21371..21503
Query Match 18.5%; Score 35.2; DB 9; Length 37731;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 6 aaagagagagcctgaactctctcttgacatcttaagatgtaagaagaataacttttga 65
DB 10522 AAAAATTACCGGACATGAGGTGTGCTGTAGTCTACTCTGAGAGGCTGAGTGGGA 10581
QY 66 gggttagaagacatgaggaatggtgttcaacttgctgcacaggaacccagctct 125
DB 10582 GAATTCACCTGAGCGCCGGGACATGAGAGCTGACAGCTGAGATGACACCTGCCTCC 10641
QY 126 tcacttgaaacgaatcacgcctcttgaaagatcatccctaagaagagaagctac 185
DB 10642 AGCCTAGGCTACAGACCAAAACCTGTTCAAAACAAACAAACAAACAAACAA 10701
QY 186 taata 189
DB 10702 AAAA 10705
RESULT 13
AC06694/c

```

LOCUS	AC066694	171902 bp	DNA	linear	HTG 08-FEB-2002							
DEFINITION	Homo sapiens chromosome 2 clone RP11-655A7, WORKING DRAFT SEQUENCE.											
ACCESSION	1 uncloned pieces.											
VERSION	AC066694.6	GI:18640720										
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFTN.											
SOURCE	human.											
ORGANISM	Homo sapiens											
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.											
AUTHORS	1 (bases 1 to 171902)											
TITLE	Waterston, R.H.											
JOURNAL	The sequence of Homo sapiens clone											
REFERENCE	unpublished											
AUTHORS	2 (bases 1 to 171902)											
TITLE	Waterston, R.H.											
JOURNAL	Direct Submission											
COMMENT	Submitted (25-APR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA											
COMMENT	On Feb 8, 2002 this sequence version replaced gi:18482349.											

```
Repeat_region complement(3434, .3738)
/rpt_family="AluY"
Repeat_region complement(3742, .3973)
/rpt_family="LIM4"
Repeat_region 4031, .4343
/rpt_family="AluY"
Repeat_region 4344, .4369
/rpt_family="AT-rich"
misc_feature 4578, .4762
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: good, score: 52.000"
Repeat_region complement(4845, .4878)
/rpt_family="AT-rich"
Repeat_region complement(4946, .5052)
/rpt_family="MER64"
Repeat_region complement(5091, .5161)
/rpt_family="MER64B"
Repeat_region 5163, .5461
/rpt_family="AluSc"
Repeat_region complement(5699, .6017)
/rpt_family="MER64"
Repeat_region 6396, .6492
/rpt_family="MIR"
Repeat_region 6588, .6619
/rpt_family="AT-rich"
Repeat_region complement(6655, .6956)
/rpt_family="AluSq"
misc_feature complement(6946, .7073)
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: excellent, score: 75.000"
Repeat_region 7220, .7266
/rpt_family="LI"
Repeat_region 7267, .7410
/rpt_family="BC200"
Repeat_region 7411, .7464
/rpt_family="LI"
Repeat_region 7467, .7763
/rpt_family="AluJo"
Repeat_region 7764, .7860
/rpt_family="(GAAA)n"
Repeat_region 7862, .8025
/rpt_family="LI"
Repeat_region complement(8027, .8482)
/rpt_family="MER74"
Repeat_region 8520, .8821
/rpt_family="AluSx"
Repeat_region complement(8838, .9815)
/rpt_family="LFR5"
Repeat_region complement(9873, .9972)
/rpt_family="(CA)n"
Repeat_region complement(10307, .10723)
/rpt_family="MLT1B"
Repeat_region complement(10935, .10967)
/rpt_family="(CA)n"
Repeat_region complement(11309, .11400)
/rpt_family="(CA)n"
misc_feature complement(11693, .11874)
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: good, score: 59.000"
Repeat_region 11966, .12160
/rpt_family="AluCb"
Repeat_region 12272, .12292
/rpt_family="AT-rich"
Repeat_region complement(12301, .12354)
/rpt_family="LIMB3"
Repeat_region 12451, .12779
/rpt_family="MER7A"
Repeat_region complement(12782, .13078)
/rpt_family="AluY"
Repeat_region complement(13088, .13765)
/rpt_family="LIMB3"
Repeat_region complement(13984, .14278)
/rpt_family="AluSx"
```

```
Repeat_region 14710, .15013
/rpt_family="AluSx"
Repeat_region 15016, .15039
/rpt_family="(CAA)n"
Repeat_region complement(15740, .15809)
/rpt_family="MER64"
Repeat_region 15906, .16202
/rpt_family="AluSx"
Repeat_region complement(16482, .16563)
/rpt_family="MER74"
Repeat_region complement(16787, .16902)
/rpt_family="(CA)n"
Repeat_region complement(17091, .17426)
/rpt_family="MLT1B"
Repeat_region complement(17590, .17630)
/rpt_family="(CA)n"
Repeat_region 17882, .17942
/rpt_family="MIR"
misc_feature complement(17941, .18120)
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: excellent, score: 75.000"
Repeat_region 18203, .18372
/rpt_family="AluSp"
Repeat_region 18392, .18690
/rpt_family="AluSx"
Repeat_region complement(18830, .18870)
/rpt_family="AT-rich"
Repeat_region complement(18955, .18980)
/rpt_family="AT-rich"
Repeat_region 19030, .19349
/rpt_family="MER7A"
Repeat_region 19343, .19415
/rpt_family="(TA)n"
Repeat_region complement(19417, .19701)
/rpt_family="AluSx"
Repeat_region complement(19703, .19842)
/rpt_family="(TA)n"
Repeat_region 19917, .20056
/rpt_family="(TA)n"
Repeat_region complement(20066, .20210)
/rpt_family="AluSp/g"
Repeat_region complement(20220, .20534)
/rpt_family="LIMB3"
Repeat_region complement(20534, .20862)
/rpt_family="LIPAI6"
Repeat_region complement(20863, .21162)
/rpt_family="AluSp"
Repeat_region complement(21175, .21527)
/rpt_family="LIPAI6"
Repeat_region complement(21574, .21847)
/rpt_family="AluY"
Repeat_region complement(21848, .22139)
/rpt_family="LIMB3"
Repeat_region 22346, .22366
/rpt_family="AT-rich"
Repeat_region complement(22619, .22711)
/rpt_family="LIMB3"
Repeat_region complement(23083, .23169)
/rpt_family="MER64B"
Repeat_region 23170, .23296
/rpt_family="TIGER1"
```

Query Match 18.4% Score 35; DB 9; Length 178367;
Best Local Similarity 55.3%; Pred. No. 2.1;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 47 taagaataactcttggagggttagaatacaaggagacatggtgttcaacttgcctc 106
|||||
DB 51135 TCAGCAAGCTGAAGTGGAGGATCACTTGAGCTGAGAGATCGCGGTGCAGTAAGCTCT 51076
caaggacacgcgcagcttcaacttggaaacagaatacaagccttgcgaagatcatccc 166
|||||
QY 107 caaggacacgcgcagcttcaacttggaaacagaatacaagccttgcgaagatcatccc 166
|||||
DB 51075 GATGGACCACTGCATTCTCGCTGGGCAACAGAGTAAGACTTGTCTCAAAAAAATCA 51016


```

OY      167 taa 169
Db      51015 TAA 51013

RESULT  15
AC073986/c
LOCUS   164368 bp DNA linear HTG 13-AUG-2000
DEFINITION Homo sapiens chromosome 5 clone RP11-427N4, WORKING DRAFT SEQUENCE,
AC073986
25 unordered pieces.
ACCESSION AC073986.2 GI:9795960
VERSION   AC073986
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 164368)
AUTHORS   Waterston,R.H.
TITLE     The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 164368)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (08-JUL-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
COMMENT   On Aug 11, 2000 this sequence version replaced gi:8980003.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0427N04
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150398 bases at least Q40
Consensus quality: 154507 bases at least Q30
Consensus quality: 156376 bases at least Q20
Insert size: 156000; agarose-fp
Insert size: 161968; sum-of-contigs
Quality coverage: 3.75 in Q20 bases; sum-of-contigs
Quality coverage: 3.82 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1598: contig of 1598 bp in length
* 1599 1698: gap of unknown length
* 1699 3738: contig of 2040 bp in length
* 3739 3838: gap of unknown length
* 3839 4963: contig of 1125 bp in length
* 4964 5063: gap of unknown length
* 5064 7067: contig of 2004 bp in length
* 7068 7167: gap of unknown length
* 7168 9526: contig of 2359 bp in length
* 9527 12145: contig of 2519 bp in length
* 12146 15270: contig of 3025 bp in length
* 15271 15370: gap of unknown length
* 15371 17747: contig of 2377 bp in length

```

```

* 17748 17847: gap of unknown length
* 17848 21615: contig of 3768 bp in length
* 21616 21715: gap of unknown length
* 21716 25342: contig of 3627 bp in length
* 25343 25442: gap of unknown length
* 25443 29469: contig of 4027 bp in length
* 29470 29569: gap of unknown length
* 29570 34443: contig of 4874 bp in length
* 34444 39595: contig of 5052 bp in length
* 39596 39695: gap of unknown length
* 39696 45834: contig of 6139 bp in length
* 45835 45934: gap of unknown length
* 45935 52225: contig of 6291 bp in length
* 52226 52325: gap of unknown length
* 52326 62588: contig of 10263 bp in length
* 62589 62688: gap of unknown length
* 62689 72677: contig of 9989 bp in length
* 72678 72777: gap of unknown length
* 72778 83043: contig of 10266 bp in length
* 83044 83143: gap of unknown length
* 83144 92619: contig of 9476 bp in length
* 92620 92719: gap of unknown length
* 92720 102437: contig of 9718 bp in length
* 102438 102537: gap of unknown length
* 102538 113400: contig of 10863 bp in length
* 113401 113500: gap of unknown length
* 113501 123821: contig of 10321 bp in length
* 123822 123921: gap of unknown length
* 123922 135948: contig of 12027 bp in length
* 135949 136048: gap of unknown length
* 136049 150259: contig of 14211 bp in length
* 150260 150359: gap of unknown length
* 150360 164368: contig of 14009 bp in length.
Location/Qualifiers
1. 164368
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="5"
   /clone="RP11-427N4"
1. 1598
   /note="assembly_name:Contig5"
1699. 3738
   /note="assembly_name:Contig6"
3839. 4963
   /note="assembly_name:Contig7"
clone_end:T7
vector_side:left
5064. 7067
   /note="assembly_name:Contig8"
7168. 9526
   /note="assembly_name:Contig9"
9627. 12145
   /note="assembly_name:Contig10"
12246. 15270
   /note="assembly_name:Contig11"
15371. 17747
   /note="assembly_name:Contig12"
17848. 21615
   /note="assembly_name:Contig13"
21716. 25342
   /note="assembly_name:Contig14"
25443. 29469
   /note="assembly_name:Contig15"
29570. 34443
   /note="assembly_name:Contig16"
34544. 39595
   /note="assembly_name:Contig17"
39696. 45834
   /note="assembly_name:Contig18"
45935. 52225
   /note="assembly_name:Contig19"
52326. 62588

```

```

BASE COUNT      49390 a 32485 c 32356 g 47732 t 2405 others
ORIGIN
misc_feature     /note="assembly_name:Contig20"
62689..72677
|note="assembly_name:Contig21"
72778..83043
|note="assembly_name:Contig22"
83144..92619
|note="assembly_name:Contig23"
92730..102437
|note="assembly_name:Contig24"
102538..113400
|note="assembly_name:Contig25"
113501..123821
|note="assembly_name:Contig26
clone_end:SP6
vector_side:left"
123922..135948
|note="assembly_name:Contig27"
136049..150259
|note="assembly_name:Contig28"
150360..164368
|note="assembly_name:Contig29"

```

Query Match	Similarity	18.1%	Score 34.4	DB 2	Length 164368
Best Local	Similarity	53.8%	Pred. No. 3.2		
Matches	71	Conservative	0	Mismatches	61
				Indels	0
				Gaps	0
QY	45	gtgcgaanaatacctcttggagggttagaagatcaggggacatggtgtgttcacattgct	104		
DB	81171	TCGCCATTTTCCCTTCCTTCACAGGTGGGAGAGAAATGTCAGGGGTTCACATTTTAA	81112		
QY	105	gccacggaacacccgcagctcttcacttggaaacagaatcaatcgcttggaaatgacatc	164		
DB	81111	GCAGCAGAGAGCTGCCACCTCGGAGGTGAGGAAAAAACCATGATGATGAGAGAACAA	81052		
QY	165	cctaagcaggag	176		
DB	81051	AAAAAGCATGAG	81040		

Search completed: September 26, 2002, 04:23:02
Job time: 8947 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 03:03:33 ; Search time 475.38 Seconds
(without alignments)
686.217 Million cell updates/sec

Title: US-09-874-390-1_COPY_246_435
Perfect score: 190
Sequence: 1 atattaagagagcctgaa.....cagagagaagactactaaag 190

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

1:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	190	100.0	267	21 AAA96572 Noncoding region o
2	190	100.0	2108	21 AAA96570 A core 2 beta-1,6-
3	190	100.0	2128	22 AAC62134 Nucleotide sequenc
4	190	100.0	2229	21 AAC99109 Human pancreatic c
5	190	100.0	2236	22 AAH34463 Human colon cancer
6	190	100.0	2319	21 AAA48623 Human C2/gnt CDNA
7	31.8	16.7	2347	22 AAK94836 Human full-length
8	31.2	16.4	2448	23 AAS69345 DNA encoding novel
9	31.2	16.4	2448	23 AAS69947 DNA encoding novel

10	31.2	16.4	2448	23 AAS72342 DNA encoding novel
11	31.2	16.4	3204	23 AAS72347 DNA encoding novel
12	31.2	16.4	3652	22 AAS99642 Human protein enco
13	31.2	16.4	4422	21 AAA09310 Human cancer assoc
14	30.6	16.1	2936	23 AAS7231 DNA encoding novel
15	30.4	16.0	2582	21 AAC54273 Arabidopsis thalla
16	30.2	15.9	15812	22 ABA18253 Human nervous syst
17	30	15.8	406	22 AA190547 Human polynucleot
18	29.6	15.6	12482	21 AAA50353 Human Goodpasture
19	29.4	15.5	1555	21 AA251314 S. cerevisiae delt
20	29.2	15.4	560	22 AAH10006 Human cDNA clone (
21	29	15.3	6616	22 AAB61014 Human immune/haema
22	28.8	15.2	556	22 ABA61053 Human foetal liver
23	28.8	15.2	556	22 AAK09346 Human brain expres
24	28.8	15.2	556	22 AAK35235 Human bone marrow
25	28.8	15.2	556	22 AAK10951 Probe #9637 used t
26	28.8	15.2	2155	21 AA251101 A. thaliana MAP k1
27	28.8	15.2	2235	23 AAS72368 DNA encoding novel
28	28.8	15.2	2249	22 AAK70693 Human immune/haema
29	28.8	15.2	2249	22 AAK70694 Human immune/haema
30	28.8	15.2	3769	23 AAS72369 DNA encoding novel
31	28.8	15.2	9049	22 AAK72210 Human immune/haema
32	28.8	15.2	9049	22 AAK72211 Human immune/haema
33	28.8	15.2	9049	22 AAK77870 Human immune/haema
34	28.8	15.2	9049	22 AAK77871 Human immune/haema
35	28.6	15.1	10552	22 ABA19272 Human nervous syst
36	28.6	15.1	14041	22 AAH48024 Internal control B
37	28.4	14.9	4444	23 ABL05116 Drosophila melanog
38	28.2	14.8	509	22 AAH36548 Human colon cancer
39	28.2	14.8	997	22 AAK70167 Human immune/haema
40	28.2	14.8	997	22 AAK81669 Human immune/haema
41	28.2	14.8	997	22 AAK82635 Human immune/haema
42	28.2	14.8	1785	22 ABA19406 Human nervous syst
43	28.2	14.8	4285	22 AAK70169 Human immune/haema
44	28.2	14.8	4285	22 AAK74894 Human immune/haema
45	28.2	14.8	4285	22 AAK82636 Human immune/haema

ALIGNMENTS

RESULT	1
AAA96572	AAA96572 standard; DNA; 267 BP.
ID	AAA96572 standard; DNA; 267 BP.
XX	XX
AC	AAA96572;
XX	XX
DT	08-FEB-2001 (first entry)
XX	XX
DE	Noncoding region of core 2 beta-1,6-N-acetylglucosaminyltransferase gene.
XX	XX
KW	Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b G1CNAc-T;
KW	cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW	rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW	septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW	platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW	clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW	diverticulitis; ulcerative colitis; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	CA2296936-A1.
XX	XX
PD	03-AUG-2000.
XX	XX
PF	03-FEB-2000; 2000CA-2296936.
XX	XX
PR	03-FEB-1999; 99US-0118674.
XX	XX
PA	(GLYC-) GLYCOSTEIN INC.
XX	XX
PI	Korczak B, Lew A;
XX	XX

DR WPI: 2000-594746/57.
XX
XX New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PR compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
XX
PS Disclosure: Page 53; 66pp; English.
XX
XX The present sequence represents a noncoding region from a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) gene.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 267 BP; 66 A; 67 C; 60 G; 74 T; 0 other;

Query Match 100.0%; Score 190; DB 21; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.6e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atattaagagagagcctgaactgtctccttgagacatctatgaatgtcagaataactt 60
DB 18 atattaagagagagcctgaactgtctccttgagacatctatgaatgtcagaataactt 77

OY 61 ttggaagggttagaagatcaggggacatggttgcacatttgcgcaggaacaccgcc 120
DB 78 ttggaagggttagaagatcaggggacatggttgcacatttgcgcaggaacaccgcc 137

OY 121 agcttcacttggaaacagaatcacgccttggaaagagatcaccctaagcaggaagaa 180
DB 138 agcttcacttggaaacagaatcacgccttggaaagagatcaccctaagcaggaagaa 197

OY 181 gctactaaag 190
DB 198 gctactaaag 207

RESULT 2
AAA96570
ID AAA96570 standard; DNA; 2108 BP.
XX
XX AAA96570;
AC
XX
DT 08-FEB-2001 (first entry)
XX
XX
DE A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.
XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase: core2b GlcNAc-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.
XX
XX Homo sapiens.
OS
XX
XX CA2296936-A1.
PN
XX
XX 03-AUG-2000.
PD

XX
XX 03-FEB-2000; 2000CA-2296936.
PF
XX
XX 03-FEB-1999; 99US-0118674.
PR
XX
XX (GLYC-) GLYCODESIGN INC.
PA
XX
XX Korczak B, Lew A;
PI
XX
XX WPI: 2000-594746/57.
DR
XX
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PR compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
XX
PS Claim 4; Page 51-52; 66pp; English.
XX
XX The present sequence encodes a partial human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;

Query Match 100.0%; Score 190; DB 21; Length 2108;
Best Local Similarity 100.0%; Pred. No. 4e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atattaagagagagcctgaactgtctccttgagacatctatgaatgtcagaataactt 60
DB 18 atattaagagagagcctgaactgtctccttgagacatctatgaatgtcagaataactt 77

OY 61 ttggaagggttagaagatcaggggacatggttgcacatttgcgcaggaacaccgcc 120
DB 78 ttggaagggttagaagatcaggggacatggttgcacatttgcgcaggaacaccgcc 137

OY 121 agcttcacttggaaacagaatcacgccttggaaagagatcaccctaagcaggaagaa 180
DB 138 agcttcacttggaaacagaatcacgccttggaaagagatcaccctaagcaggaagaa 197

OY 181 gctactaaag 190
DB 198 gctactaaag 207

RESULT 3
AAC62134
ID AAC62134 standard; DNA; 2128 BP.
XX
XX AAC62134;
AC
XX
DT 06-MAR-2001 (first entry)
XX
XX
DE Nucleotide sequence of beta-1,6-N-acetylglucosaminyltransferase.
XX
XX Human; beta-1,6-N-acetylglucosaminyltransferase; C2GNT-M; inflammation;
KW membrane protein; branched sialyl lex; L-selection; immune reaction;
KW inflammation; tissue rejection; tumour metastasis; ss.
XX

```
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 354..1670
XX FT /tag= a
XX FT /product= "beta-1-6-N-acetylglucosaminyltransferase"
XX FT 2100..2105
XX FT /tag= b
XX
XX US6136580-A.
XX
XX 24-OCT-2000.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX (BURN-) BURNHAM INST.
XX
XX Fukuda M, Yeh J;
XX
XX WPI: 2001-040238/05.
XX P-PSDB; AAB30518.
XX
XX New C3Gnt-M polypeptides having core 2, core 4 and I branching
XX beta-1-6-N-acetylglucosaminyltransferase activities for preparing
XX reagents useful for diagnosing, preventing or treating inflammation or
XX tumor metastasis
XX
XX Example 1; Fig 4; 25pp; English.
XX
XX The present sequence encodes a human
XX beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
XX and I branching activities. It is designated C3Gnt-M. C3Gnt-M is a
XX membrane protein that is predominantly expressed in colon, small
XX intestine, trachea, stomach and thyroid, as well as in certain cancer
XX cell lines. C3Gnt-M polypeptides may be used to prepare molecules having
XX highly branched sialyl lex and I-selectins, which may be subsequently
XX used to modulate immune reactions, e.g. inflammation and tissue
XX rejection, and to prevent or inhibit tumor metastasis.
XX
XX Sequence 2128 BP; 569 A; 477 C; 526 G; 556 T; 0 other;
S0

Query Match 100.0%; Score 190; DB 22; Length 2128;
Best Local Similarity 100.0%; Pred. NO. 4e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataacctt 60
DB 1104 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataacctt 163
OY 61 ttgaggggttgaagatcagggagacatggttctacatttgctccacggaaacggcc 120
DB 164 ttgaggggttgaagatcagggagacatggttctacatttgctccacggaaacggcc 223
OY 121 agcttcacttgaagaacgaatcacgccttgaagagatcatcccttaagcagagagaa 180
DB 224 agcttcacttgaagaacgaatcacgccttgaagagatcatcccttaagcagagagaa 283
OY 181 gctactaaag 190
DB 284 gctactaaag 293

RESULT 4
AAC99109
ID AAC99109 standard; cDNA; 2229 BP.
XX
XX AAC99109;
XX
XX 09-MAR-2001 (first entry)
XX
```

```
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
XX
XX Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
XX detection; diagnosis; identification; cytostatic; neuroprotective;
XX nocotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
XX antinflammatory; cardiant; gene therapy; chromosome mapping;
XX linkage analysis; tissue identification; tissue typing; forensic;
XX neural; immune system; muscular; reproductive; gastrointestinal;
XX pulmonary; cardiovascular; renal; proliferative; ss.
XX
XX Homo sapiens.
XX
XX WO200055320-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-579444/54.
XX P-PSDB; AAB54344.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 1; Page 759-760; 1379pp; English.
XX
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX proteins, called pancreatic cancer antigens, given in AAB54008 to
XX AAB54466. The human pancreatic cancer antigens have cytostatic,
XX neuroprotective, nocotropic, immunomodulatory, relaxant, contraceptive,
XX gynaecological, cardiant and antinflammatory activities, and can be used
XX in gene therapy. The polynucleotide and proteins can be used for
XX preventing, treating, or ameliorating a medical condition or in assays
XX for diagnosing a pathological condition or a susceptibility to one in a
XX subject. Binding partners to the proteins and the activity of the
XX proteins can be identified. The pancreatic cancer antigens can be used to
XX detect, treat or prevent pancreatic disorders, especially cancer.
XX Agonists and antagonists to the antigens can be screened for. The
XX pancreatic cancer antigen polynucleotides can be used to design nucleic
XX acid hybridisation probes that can be used in chromosome mapping, linkage
XX analysis, tissue identification and/or typing and a variety of forensic
XX and diagnostic methods. The proteins can be used to generate antibodies
XX which are used to purify, detect and target the polypeptides, including
XX both in vivo and in vitro diagnostic and therapeutic methods. The
XX proteins can be used to treat or prevent neural, immune system, muscular,
XX reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;
S0

Query Match 100.0%; Score 190; DB 21; Length 2229;
Best Local Similarity 100.0%; Pred. NO. 4.1e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataacctt 60
DB 177 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataacctt 236
OY 61 ttgaggggttgaagatcagggagacatggttctacatttgctccacggaaacggcc 120
DB 237 ttgaggggttgaagatcagggagacatggttctacatttgctccacggaaacggcc 296
OY 121 agcttcacttgaagaacgaatcacgccttgaagagatcatcccttaagcagagagaa 180
DB 284 gctactaaag 293
```

```

Db      297 agcttcacttggaaacagaatcacgccttgtgaaagatcattccctaagcaggagagaa 356
Oy      181 gctactaaag 190
        |||||||
Db      357 gctactaaag 366

RESULT  5
AAH34463
ID      AAH34463 standard; cDNA; 2236 BP.
XX
AC      AAH34463;
XX
DT      03-SEP-2001 (first entry)
XX
DE      Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
XX
KM      Human: colon cancer; colon cancer antigen; diagnosis; detection;
XX      colorectal carcinoma; chromosome 15; ss.
XX
OS      Homo sapiens.
XX
PN      WC200122920-A2.
XX
PD      05-APR-2001.
XX
PF      28-SEP-2000; 2000MO-US26524.
XX
PR      29-SEP-1999; 99US-0157137.
PR      03-NOV-1999; 99US-0163280.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR      WPI: 2001-235357/24.
DR      P-PSDB; AAG75058.
XX
PT      Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX      useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS      Claim 1; Page 3190-3191; 9803pp; English.
XX
XX      AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX      cancer-associated nucleic acid molecules (N) and proteins (P), where
XX      the proteins are collectively known as colon cancer antigens. The colon
XX      cancer antigens have cytostatic activity and can be used in gene
XX      therapy and vaccine production. N and P may be used in the prevention,
XX      diagnosis and treatment of diseases associated with inappropriate P
XX      expression. For example, N and P may be used to treat disorders
XX      associated with decreased expression by rectifying mutations or deletions
XX      in a patient's genome that affect the activity of P by expressing P.
XX      Inactive proteins or to supplement the patient's own production of P.
XX      Additionally, N may be used to produce the colon cancer-associated PS,
XX      by inserting the nucleic acids into a host cell and culturing the cell
XX      to express the proteins. N and P can be used in the prevention, diagnosis
XX      and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX      and AAB77769 represent sequences used in the exemplification of the
XX      present invention.
XX
CC      N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX      missing at time of publication, meaning no sequences are present for
XX      SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ      Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other;

Query Match      100.0%; Score 190; DB 22; Length 2236;
Best Local Similarity 100.0%; Pred. No. 4,1e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      1 attttaaaggagagccgaactgttcttggcaccatttaagaaatgcagaataactctt 60
        |||||||
Db      184 atattaaaggagagccgtgaaccgttctcttgacacattctatgaatgycagaataactctt 243

```

```

Oy      61 ttggaagggttagaagaatcaggggacatggtgtgtcacattgtctgcacggaacaccgcc 120
        |||||||
Db      244 ttggaagggttagaagaatcaggggacatggtgtgtcacattgtctgcacggaacaccgcc 303
Oy      121 agcttcacttggaaacagaatcacgccttgtgaaagatcattccctaagcaggagagaa 180
        |||||||
Db      304 agcttcacttggaaacagaatcacgccttgtgaaagatcattccctaagcaggagagaa 363
Oy      181 gctactaaag 190
        |||||||
Db      364 gctactaaag 373

RESULT  6
AAH48623
ID      AAH48623 standard; cDNA; 2319 BP.
XX
AC      AAH48623;
XX
DT      19-SEP-2000 (first entry)
XX
DE      Human C2/4GnT cDNA.
XX
KM      Human; C2/4GnT; UDP-N-acetylglucosamine; O-glycan biosynthesis;
XX      O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
XX      CDS      496..1812
XX      FT      /*tag= a
XX      FT      /product= "C2/4GnT"
XX      FT      primer_bind 586..605
XX      FT      primer_bind 1794..1813
XX      FT      /*tag= b
XX      FT      /*tag= c
XX      FT      polyA_signal 2244..2249
XX      FT      /*tag= d
XX
PN      WO200034449-A2.
XX
PD      15-JUN-2000.
XX
XX      03-DEC-1999; 99MO-DK00677.
XX
XX      04-DEC-1998; 98DK-0001605.
XX
PA      (CLAU/) CLAUSEN H.
XX
PI      Clausen H, Schwientek T;
XX
DR      WPI: 2000-423407/36.
DR      P-PSDB; AAY94492.
XX
PT      New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
XX      probe for the detection of specified glucosaminyltransferase from
XX      other species and related organisms -
XX
PS      Claim 5; Fig 2; 47pp; English.
XX
XX      The present sequence encodes human UDP-N-acetylglucosamine:
XX      N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
XX      (C2/4GnT). The protein is the third member of the family of O-glycan
XX      beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
XX      based O-glycans on to oligosaccharides, glycoproteins and
XX      glycosphingolipids. C2/4GnT can therefore be used in the production of
XX      appropriately glycosylated glycoconjugates with particular enzymatic,
XX      immunogenic, or other biological or physical properties. The nucleotide
XX      sequence is useful as a probe for the detection of C2/4GnT from other
XX      species and related organisms and for the recombinant production of
XX      C2/4GnT polypeptide. The nucleotide sequence was identified by analysis
XX      of EST database sequence information. Oligonucleotides derived from EST

```

CC clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
CC from a human foreskin genomic pl library by 5' RACE PCR. RT-PCR was
CC performed using Coloz05 human cell line mRNA in order to produce cDNA
CC for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
CC has been implicated in tumour progression and metastasis.
XX
SQ Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other;

Query Match 100.0%; Score 190; DB 21; Length 2319;
Best Local Similarity 100.0%; Pred. No. 4,2e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atataaagagagccttaaaccttctccttgacatctatcgaatgtaagaataactt 60
Db 246 atataaagagagccttaaaccttctccttgacatctatcgaatgtaagaataactt 305
QY 61 ttggaagggttagaagatcaggagacatggtgttcacattgtcgtccagaaacacgcgc 120
Db 306 ttggaagggttagaagatcaggagacatggtgttcacattgtcgtccagaaacacgcgc 365
QY 121 agcttcaccttgaaacagatcacgcctgttgaaagatcatcccttaagcaggagaa 180
Db 366 agcttcaccttgaaacagatcacgcctgttgaaagatcatcccttaagcaggagaa 425
QY 181 gctactaaag 190
Db 426 gctactaaag 435

RESULT 7
AAK94836
ID AAK94836 standard; cDNA; 2347 BP.

XX AAK94836;

XX 06-NOV-2001 (first entry)

XX Human full-length cDNA, SEQ ID NO: 3990.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota F, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI: 2001-524255/58.

XX P-PSDB: AAM93876.

XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -

XX Claim 8; SEQ ID NO 3990; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 2347 BP; 555 A; 618 C; 751 G; 423 T; 0 other;

Query Match 16.7%; Score 31.8; DB 22; Length 2347;
Best Local Similarity 51.0%; Pred. No. 0.71;
Matches 75; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 6 aaagagagcctgaacacgttccttgagcactatgatgtcgaataatcctttga 65
Db 1867 aaagagagcctgaacacgttccttgagcactatgatgtcgaataatcctttga 1926
QY 66 gggttagaagatcaggaggacatggtgttcacattgtcgcacagaaacaccgcagct 125
Db 1927 gggttagaagatcaggaggacatggtgttcacattgtcgcacagaaacaccgcagct 1986
QY 126 tcaacttggaacagaatcacgccttgt 152
Db 1987 agcttggtcacagagtaagacctgt 2013

RESULT 8
AAS69345
ID AAS69345 standard; cDNA; 2448 BP.

XX AAS69345;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #5149.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HXSE-) HXSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: ABG05158.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -

XX Claim 1; SEQ ID NO 5149; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 2448 BP; 726 A; 481 C; 599 G; 642 T; 0 other;

Query Match 16.4%; Score 31.2; DB 23; Length 2448;
Best Local Similarity 57.0%; Pred. No. 1.2;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 9 gagagagcctgaactgttcttgccttgacatcttatgaatgcagaataacctttggaggg 68
Db 1420 gagagagcagaagaatcagtcagtcattcctctggaatgacatagacacctttggagag 1479

Qy 69 ttagaagatcagggagacatgtgttcacattgctgcga 108
Db 1480 cgagagcatcacaacagtggttattccagagaagtggtccca 1519

RESULT 11
AAS72347
ID AAS72347 standard; cDNA; 3204 BP.

AC AAS72347;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #8151.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

DR WPI, 2001-639362/73.

XX P-PSDB; ABG08160.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 1: SEQ ID NO 8151; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 3204 BP; 996 A; 572 C; 691 G; 943 T; 2 other;

Query Match 16.4%; Score 31.2; DB 23; Length 3204;
Best Local Similarity 57.0%; Pred. No. 1.3;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 9 gagagagcctgaactgttcttgccttgacatcttatgaatgcagaataacctttggaggg 68
Db 919 gagagagcagaagaatcagtcagtcattcctctggaatgacatagacacctttggagag 978

Qy 69 ttagaagatcagggagacatgtgttcacattgctgcga 108
Db 979 cgagagcatcacaacagtggttattccagagaagtggtccca 1018

RESULT 12
AAH99642/c
ID AAH99642 standard; cDNA; 3652 BP.

AC AAH99642;

DT 16-OCT-2001 (first entry)

DE Human protein encoding cDNA sequence SEQ ID NO:477.

XX

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;

KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;

KW antibacterial; endocrine; cardiac; central nervous system; vitruide;

KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;

KW antiagregant; haemostatic; vulnery; antilucer; osteopathic; eczema;

KW dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic;

KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;

KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;

KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;

KW genetic disease; haematopoietic disorder; platelet disorder; asthma;

KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;

KW allergic rhinitis; diabetes; multiple sclerosis; depression;

KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;

XX neurological disorder; ss.

XX Homo sapiens.

PN WO200153455-A2.

PD 26-JUL-2001.

PF 22-DEC-2000; 2000WO-US35017.

```

PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0486725.
PR 25-APR-2000; 2000US-0552317.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dirmanac RT;
XX WPI; 2001-457603/49.
XX P-PSDB; AAM25701.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 1; Page 549-550; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX AAM25963. The proteins can have activities based on the tissues and
XX cells they are expressed in, such as: antiinflammatory; antithrombotic;
XX antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
XX central nervous system; virucide; anti-HIV; fungicide; antitumor;
XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
XX antituber; osteopathic; dermatological; antiallergic; antischismatic;
XX antidiabetic; cyostatic; neuroprotective; antidepressant; nootropic;
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX encoding them can be used in gene therapy, antisense therapy and vaccine
XX production. The proteins and polynucleotides are useful for screening for
XX agonists or antagonists of a protein and for the treatment and diagnosis
XX of disorders associated with the activity of a protein e.g. inflammation,
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX osteoporosis, severe combined immunodeficiency, eczema, allergic
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX neurological disorders.
XX
XX SQ Sequence 3652 BP; 1079 A; 796 C; 650 G; 1127 T; 0 other;

Query Match          16.4%; Score 31.2; DB 22; Length 3652;
Best Local Similarity 57.0%; Pred. No. 1.4;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 9 gagggagcctgaacgttctcttgagacatcttgaatgacagaataacctttggagg 68
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2233 GAGGAGACAGAAAGTCATGTGCTATCTCTCGAAATGACATAGGACTTTTGGAGAG 2174
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 69 ttgaagaatcaggggacatggtgttgcacattgctgcga 108
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2173 CGAGAGCATCAACAGGTGTATTCCAGAAAGTTGTTCCCA 2134

RESULT 13
AAA09310
ID AAA09310 standard; DNA; 4422 BP.
XX
XX AAA09310;
AC
XX
XX 10-AUG-2000 (first entry)
XX
XX Human cancer associated antigen precursor DNA, clone NY-BEN-45.
XX
XX DE Human cancer associated antigen precursor; diagnosis;
XX
XX KW renal cancer; cancer associated antigen precursor; diagnosis;
XX
XX KW cyostatic; ss.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 861..3308
XX FT /*tag=
XX FT /transl_except= (pos:1473..1480, aa:Ala)

```

```

FT
XX
XX
XX WO200020587-A2.
XX
XX
XX 13-APR-2000.
XX
XX
XX 04-OCT-1999; 99WO-US22873.
XX
XX
XX 05-OCT-1998; 98US-0166300.
XX
XX 05-OCT-1998; 98US-0166350.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Obata Y, Gout I, Tureci O, Sahin U, Pfeundschuh M, Scanlan MJ;
XX Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
XX
XX WPI; 2000-303774/26.
XX P-PSDB; AAY92338.
XX
XX Preventing, diagnosing and/or treating disorders associated with
XX abnormal expression of human cancer associated antigens
XX
XX Claim 57; Page 80-81; 121pp; English.
XX
XX AAA09310-20 are novel genes isolated by SEREX screening from a renal
XX cancer cell line 1973/10.4. The genes encode cancer associated antigen
XX precursors. These gene products are useful in methods for preventing,
XX diagnosing and/or treating disorders, especially cancer, associated with
XX abnormal expression of human cancer associated antigens. The method
XX comprises contacting a sample from a subject with an agent that
XX specifically binds to the nucleic acid molecule or expression product
XX (or fragment) complexed with a human leukocyte antigen (HLA) molecule
XX and determining the interaction between the agent and the nucleic acid
XX molecule or the expression product as a determination of the disorder.
XX
XX SQ Sequence 4422 BP; 1355 A; 814 C; 978 G; 1270 T; 5 other;

Query Match          16.4%; Score 31.2; DB 21; Length 4422;
Best Local Similarity 57.0%; Pred. No. 1.5;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 9 gagggagcctgaacgttctcttgagacatcttgaatgacagaataacctttggagg 68
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2280 gagggagacagaatcattgtagctattctcttggaatgacataggacctttggagg 2339
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 69 ttgaagaatcaggggacatggtgttgcacattgctgcga 108
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2340 cgagacgatcaacagtggtttatccagaagaagtgttccca 2379

RESULT 14
AAS72731/C
ID AAS72731 standard; CDNA; 2936 BP.
XX
XX AAS72731;
AC
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #8535.
XX
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.

```

PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX P-PSDB; ABG08544.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 8535; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2936 BP; 935 A; 544 C; 664 G; 793 T; 0 other;

Query Match 16.1%; Score 30.6; DB 23; Length 2936;
Best Local Similarity 68.9%; Pred. No. 2.1;
Matches 42; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 tatlaagggagccgaactgttccttgacatctatgaatgcagaataaccttt 61
DB 1892 TTTTTCAGAGAACCCAGATCCCTGCTTGATATATTAATGTTGAATAACCTTC 1833
QY 62 t 62
DB 1832 t 1832

RESULT 15
AAC54273
ID AAC54273 standard; DNA; 2582 BP.
XX
AC AAC54273;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77298.
XX
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
PN
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135553.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

```

PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149728.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      16.0%; Score 30.4; DB 21; Length 2582;
Best Local Similarity 59.1%; Pred. No. 2.3;
Matches 52; Conservative 0; Mismatches 36; Indels 0; Gaps 0.

QY 46 gtcaaaatacctttggaggttagaatcaggagacatggttgcacatttcgt 105
    | ||||| ||| | ||| | | | | | | | | | | | | | | | | |
Db 166 gaccagaataacttgtaggaacaataagatatgatcgatgctaagaagtgcgt 1745
    | ||| ||| ||| | | |

QY 106 ccacggaacccgcacgtctcacctgg 133
    | ||| ||| ||| | | |
Db 1746 ggtcttcacaacccaaccttcttcgt 1773

```

Search completed: September 26, 2002, 03:03:36
Job time: 4291 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:22:19 : Search time 94.91 Seconds
(Without alignments)
491.733 Million cell updates/sec

Title: US-09-874-390-1_COPY_246_435

Perfect score: 190
Sequence: 1 atattaagaagagcgtctgaa.....cagagaagaactactaag 190

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/1na/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/1na/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/1na/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/1na/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	190	100.0	2128	3	US-09-233-506-1
2	27.2	14.3	720	6	5455030-14
3	27.2	14.3	84495	4	US-09-797-906-3
4	27	14.2	4411529	4	US-09-103-840A-1
5	26.4	13.9	13473	5	PCT-US96-03916-1
6	26.4	13.9	18912	5	PCT-US96-03916-59
7	26.2	13.8	2200	1	US-08-272-255-21
8	26.2	13.8	2200	5	PCT-US95-08565-21
9	26	13.7	1839	3	US-08-152-019A-44
10	26	13.7	1839	3	US-08-482-677-1
11	26	13.7	4895	3	US-09-053-866-1
12	25.6	13.5	1730	2	US-08-687-080-114
13	25.6	13.5	1948	4	US-09-360-197-11
14	25.6	13.5	2688	1	US-08-088-633-3
15	25.6	13.5	2688	1	US-08-245-756-3
16	25.6	13.5	2688	1	US-08-441-750-3
17	25.6	13.5	2688	2	US-08-441-751-3
18	25.6	13.5	2688	5	PCT-US93-02521-3
19	25.6	13.5	5893	2	US-08-687-080-44
20	25.6	13.5	8982	2	US-08-976-255-5
21	25.2	13.3	12537	2	US-08-611-280-4
22	25.2	13.3	12537	4	US-09-195-940-4
23	25	13.2	1676	4	US-09-143-954-3
24	25	13.2	1702	4	US-09-143-954-1
25	24.8	13.1	636	4	US-09-328-111-226
26	24.8	13.1	969	4	US-08-936-165A-140
27	24.8	13.1	5408	1	US-08-471-058-20

28	24.8	13.1	5408	3	US-08-471-057-20	Sequence 20, Appl
29	24.6	12.9	1539	2	US-08-828-596-1	Sequence 1, Appl
30	24.6	12.9	1666	4	US-09-360-197-5	Sequence 5, Appl
31	24.6	12.9	1920	4	US-09-534-638-6	Sequence 6, Appl
32	24.6	12.9	2264	4	US-09-126-109-9	Sequence 9, Appl
33	24.6	12.9	2371	2	US-08-343-443B-1	Sequence 1, Appl
34	24.6	12.9	6156	4	US-08-891-640-1	Sequence 1, Appl
35	24.6	12.9	6156	4	US-09-723-535-3	Sequence 3, Appl
36	24.6	12.9	9840	4	US-09-534-638-1	Sequence 1, Appl
37	24.6	12.9	246240	2	US-08-724-394A-20	Sequence 20, Appl
38	24.6	12.9	246240	2	US-08-724-394A-21	Sequence 21, Appl
39	24.6	12.9	246240	2	US-08-724-394A-22	Sequence 22, Appl
40	24.4	12.8	243	4	US-08-905-223-63	Sequence 63, Appl
41	24.4	12.8	599	2	US-08-557-128-9	Sequence 9, Appl
42	24.4	12.8	2075	1	US-08-238-163-3	Sequence 3, Appl
43	24.4	12.8	35100	2	US-08-770-379-17	Sequence 17, Appl
44	24.4	12.8	35100	4	US-08-757-669A-17	Sequence 17, Appl
45	24.4	12.8	35100	4	US-09-230-371A-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-233-506-1
Sequence 1, Application US/09233506
Patent No. 6136580
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
APPLICANT: Yeh, Jiumn Chern
TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
FILE REFERENCE: P-LJ 3415
CURRENT APPLICATION NUMBER: US/09/233,506
CURRENT FILING DATE: 1999-01-19
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2128
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (354)..(1670)
US-09-233-506-1

Query Match 100.0%; Score 190; DB 3; Length 2128;
Best Local Similarity 100.0%; Pred. No. 6e-60;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagcgtctgaaactgttctcttgagatcttatagaatgtaagaataactt 60
|||||
Db 104 atattaagaagagcgtctgaaactgttctcttgagatcttatagaatgtaagaataactt 163

QY 61 ttgagaggttgaagatcagagggacatgttctacatttgcctccaggaacccgc 120
|||||
Db 164 ttgagaggttgaagatcagagggacatgttctacatttgcctccaggaacccgc 223

QY 121 agcttcaacttgaagaacgaatcagccttcttgaagagatcatccctaagcagaagaa 180
|||||
Db 224 agcttcaacttgaagaacgaatcagccttcttgaagagatcatccctaagcagaagaa 283

QY 181 gctactaag 190
|||||
Db 284 gctactaag 293

RESULT 2
5455030-14
Patent No. 5455030
APPLICANT: LADNER, ROBERT C.; BIRD, ROBERT E.; HARDMAN, KARL
TITLE OF INVENTION: IMMUNOTHERAPY USING SINGLE CHAIN

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A

RESULT 5
 PCT-US96-03916-1/c
 : Sequence 1, Application PC/TUS9603916
 GENERAL INFORMATION:
 APPLICANT: Wild, Martha A.
 APPLICANT: Cochran, Mark D.
 TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
 TITLE OF INVENTION: AND USES THEREOF
 NUMBER OF SEQUENCES: 72
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue Of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/03916
 FILING DATE: 23-MAR-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/126,597
 FILING DATE: 24-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28, 678
 REFERENCE/DOCKET NUMBER: 39116-A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 13473 base pairs
 TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1059..2489
FEATURE:
NAME/KEY: CDS
LOCATION: 2575..4107
FEATURE:
NAME/KEY: CDS
LOCATION: 4113..4445
FEATURE:
NAME/KEY: CDS
LOCATION: 4609..5487
FEATURE:
NAME/KEY: CDS
LOCATION: 5697..8654
FEATURE:
NAME/KEY: CDS
LOCATION: 9874..10962
FEATURE:
NAME/KEY: CDS
LOCATION: 11159..12658
FEATURE:
NAME/KEY: CDS
LOCATION: 12665..13447
PCT-US96-03916-1

Query Match 13.9%; Score 26.4; DB 5; Length 13473;
Best Local Similarity 61.8%; Pred. No. 14;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 6 aaagaggaagcctgaactgtctccttgagacatctatgaatgacgaataaccttttga 65
Db 7923 ATAGGGGCGCTGCGACATCTCGTCGACATCTCATGGAATTGGAACACAGTAGGAA 7864
QY 66 gggttaga 73
Db 7863 ACGATAAA 7856

RESULT 6
PCT-US96-03916-59/C
Sequence 59, Application PC/TUS9603916
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03916
FILING DATE: 23-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,597
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39116-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 18912 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 697..1533
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1900..2784)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (2916..3605)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 3694..5124
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 5210..7081
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 7245..8123
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 8333..11290
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 11098..12402
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 12510..13598
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 13792..15291
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 15298..16080
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 16129..17013
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (17380..18216)
OTHER INFORMATION:
PCT-US96-03916-59

Query Match 13.9%; Score 26.4; DB 5; Length 18912;
Best Local Similarity 61.8%; Pred. No. 17;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 6 aaagagagccttgaacgttcttgccttgacatctatgaatgaagaacacgtttgga 65
10559 ATAGGGCGCTCTGCGACTCTGCGCGTGCACATCTCATGGAATTTGAAACAAAGTAGGAA 10500
Db 66 ggggttaga 73
10499 ACGATAAA 10492

RESULT 7

US-08-272-255-21/c
Sequence 21, Application US/08272255
Patent No. 5824859
GENERAL INFORMATION:
APPLICANT: Cashmore, Anthony R.
APPLICANT: Ahmad, Margaret
APPLICANT: Lin, Chentao
TITLE OF INVENTION: Blue light photoreceptors and Methods of
TITLE OF INVENTION: Using the Same
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/272,255
FILING DATE: 08-JUL-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: UPN-1795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-272-255-21

Query Match 13.8%; Score 26.2; DB 1; Length 2200;
Best Local Similarity 60.6%; Pred. No. 6.8;
Matches 43; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 103 ctgcacaggaacacccgcttcttgccttggaacagacagacgttgaagagatca 162
Db 362 CTTCACGAGTAACGTCTCTTCTTTCAGAGACCAATGAGAGAGAGAAACAGATCC 303
Qy 163 tccctaagcag 173
Db 302 TTGCTGAGCAG 292

RESULT 8

PCT-US95-08565-21/c
Sequence 21, Application PC/TUS9508565
GENERAL INFORMATION:
APPLICANT: Cashmore, Anthony R.
APPLICANT: Ahmad, Margaret

APPLICANT: Lin, Chentao
TITLE OF INVENTION: Blue light photoreceptors and Methods of
TITLE OF INVENTION: Using the Same
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08565
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/272,255
FILING DATE: 08-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: UPN-1795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2200 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US95-08565-21

Query Match 13.8%; Score 26.2; DB 5; Length 2200;
Best Local Similarity 60.6%; Pred. No. 6.8;
Matches 43; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 103 ctgcacaggaacacccgcttcttgccttggaacagacagacgttgaagagatca 162
Db 362 CTTCACGAGTAACGTCTCTTCTTTCAGAGACCAATGAGAGAGAGAAACAGATCC 303
Qy 163 tccctaagcag 173
Db 302 TTGCTGAGCAG 292

RESULT 9

US-08-152-019A-44/c
Sequence 44, Application US/08152019A
Patent No. 5565331
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: Serafini, Tito
APPLICANT: Kennedy, Timothy
APPLICANT: Placzek, Marysia
APPLICANT: Jessell, Thomas
TITLE OF INVENTION: NEURAL AXON OUTGROWTH MODULATORS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

REGISTRATION NUMBER: 36,62/
REFERENCE/DOCKET NUMBER: UC93-300-4

```

; TOPOLOGY: Linear
FEATURE:

```



```

?      FILING DATE:
?      CLASSIFICATION: 435
?      PRIOR APPLICATION DATA:
?      APPLICATION NUMBER: 07/678,916
?      FILING DATE: 01-APR-1991
?      ATTORNEY/AGENT INFORMATION:
?      NAME: Reiter, Stephen E
?      REGISTRATION NUMBER: 31192
?      REFERENCE/DOCKET NUMBER: 50848
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (619)552-1311
?      TELEFAX: (619)552-0095
?      TELEX: 20 6566 PATIAM CGO
?      INFORMATION FOR SEQ ID NO: 3:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 2688 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: unknown
?      TOPOLOGY: unknown
?      MOLECULE TYPE: cDNA
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 643..1431
?      FEATURE:
?      NAME/KEY: mat_peptide
?      LOCATION: 643..1431
?      OS-08-088-633-3

```

Query Match	13.5%	Score 25.6	DB 1	Length 2688
Best Local Similarity	66.1%	Pred. No. 12		
Matches	37	Conservative	0	Mismatches 19; Indels 0; Gaps 0.
Oy	20	aactgtcccttgacatcttatgaatgcagaataacctttagagggtgaaag	75	
Db	1780	AGCCGTTCAATACCATTCTTGAAGTTCAGTCACATCATTTTAGGAGAAGGTGGAA	1725	

RESULT 15
 US-08-245-756-3/C
 Sequence 3, Application US/08245756
 Patent No. 554112
 GENERAL INFORMATION:
 APPLICANT: Gleeson, Martin A
 APPLICANT: Howard, Bradley D
 TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
 TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
 STREET: 444 So. Flower Street, Suite 2000
 CITY: Los Angeles
 STATE: CA
 COUNTRY: USA
 ZIP: 90071
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/245,756
 FILING DATE: 16-May-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/088,633
 FILING DATE: 06-JULY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/678,916
 FILING DATE: 01-APR-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Seidman, Stephanie
 REGISTRATION NUMBER: 33,779

```

1 . REFERENCE/DOCKET NUMBER: 9763
2 TELECOMMUNICATION INFORMATION:
3 TELEPHONE: 619-546-4737
4 TELEFAX: 619-546-9392
5 INFORMATION FOR SEQ ID NO: 3:
6 SEQUENCE CHARACTERISTICS:
7     LENGTH: 2688 base pairs
8     TYPE: nucleic acid
9     STRANDEDNESS: unknown
10    TOPOLOGY: unknown
11    MOLECULE TYPE: cDNA
12    FEATURE:
13        NAME/KEY: CDS
14        LOCATION: 643..1431
15    FEATURE:
16        NAME/KEY: mat_peptide
17        LOCATION: 643..1431
18
19 US-08-245-756-3

```

	Query Match	13.5%	Score 25.6;	DB 1,	length 2688;
	Best Local Similarity	66.1%;	Pred. No. 12;		
Matches	37;	Conservative	0;	Mismatches	19; Incls 0; Gaps 0;
Oy	20	aacgttcctcggacaatcttatgatgtaagaaataactcttttgagggttgaag	75		
Db	1780	AGCCCTTAATACCACTTTTCTAAAGTCGACGCAATATTACTTTTGACGAAGGTGAAG	1725		

Search completed: September 26, 2002, 04:25:28
Job time: 9058 sec

• • • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 02:54:56 ; Search time 3532.1 Seconds
(without alignments)
7199.191 Million cell updates/sec

Title: US-09-874-390-1_COPY_436_2319

Perfect score: 1884

Sequence: 1 gattggtctctctccacact.....ataataaacttaacaga 1884

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923.6	49.0	1862	11 AK008234	AK008234 Mus muscu
2	693.4	36.8	866	10 BG386598	BG386598 602454847
3	691.6	36.7	731	10 BG468641	BG468641 602510243
4	666	35.4	875	10 BG470202	BG470202 602533711
5	640	34.0	640	9 AA583339	AA583339 nm45f11.s
6	633.4	33.6	842	10 BE869192	BE869192 603445191
7	622.6	33.0	655	9 AM842622	AM842622 MR2-CN003
8	613.2	32.5	724	10 BG468448	BG468448 602532948
9	611	32.4	612	10 BE292814	BE292814 601105179
10	607.2	32.2	1050	10 BE867668	BE867668 601443206
11	605	32.1	671	9 A1587061	A1587061 tr55902.x
12	602.6	32.0	766	10 B1765410	B1765410 603050445
13	599.8	31.8	603	9 AM192901	AM192901 x155b12.x
14	582.6	30.9	606	9 AM662377	AM662377 h125h03.x
15	561.2	29.8	864	10 BG821420	BG821420 602724876
16	557.2	29.6	823	10 BG169857	BG169857 602324322
17	552	29.3	659	10 BE304708	BE304708 601105907

18	518	27.5	872	10 BG747491	BG747491 602704606
19	516.4	27.4	554	9 AA307800	AA307800 EST178656
20	479.4	25.4	1003	10 BG385575	BG385575 602453872
21	463.2	24.6	609	10 BF881417	BF881417 PM0-ET020
22	445.8	23.7	826	10 BG821370	BG821370 602724814
23	436	23.1	436	9 AA566218	AA566218 n18907.s
24	433	22.0	433	9 AA583146	AA583146 nm37a11.s
25	416.2	22.1	920	10 BG386247	BG386247 602455439
26	396.6	21.1	912	10 BG469141	BG469141 602511125
27	394.6	20.9	912	10 BE871910	BE871910 600447936
28	383	20.3	383	9 A1150400	A1150400 qf40h07.x
29	381.2	20.2	567	12 AZ388491	AZ388491 IM0146G03
30	377.4	20.0	420	9 AM374998	AM374998 MR0-CT006
31	375.2	19.9	965	10 BF582709	BF582709 602094124
32	373.2	19.8	390	9 A1955582	A1955582 wt58e06.x
33	368.6	19.6	702	10 BG248304	BG248304 602400340
34	362.2	19.2	411	9 AM374874	AM374874 MR0-CT006
35	362.2	19.2	472	10 B1018802	B1018802 MR4-MT025
36	358	19.0	358	9 A1832775	A1832775 at63h05.x
37	353.6	18.8	629	10 BG167918	BG167918 602340003
38	344.2	18.3	357	10 BG984134	BG984134 I15-CN006
39	343.6	18.2	438	9 AM842590	AM842590 MR2-CN003
40	333.8	17.7	882	10 BG171085	BG171085 602324130
41	317.4	16.8	361	10 BF884849	BF884849 PM0-ET020
42	317	16.8	736	10 BE617113	BE617113 601441649
43	313.4	16.6	978	10 BG171515	BG171515 602322061
44	310.6	16.5	337	9 AM887761	AM887761 PM2-OT009
45	304.4	16.2	955	10 BG328280	BG328280 602427256

ALIGNMENTS

RESULT 1

AK008234	LOCUS	AK008234	1862 bp	mRNA	linear	HTC 19-JAN-2002
AK008234	DEFINITION	Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010013H22;homolog to beta-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE, full insert sequence.				
AK008234	ACCESSION	AK008234.1	GI:12842295			
AK008234	VERSION	HTC: CAP trapper.				
AK008234	KEYWORDS	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
AK008234	SOURCE	clone:2010013H22.				
AK008234	ORGANISM	Mus musculus				
AK008234	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AK008234	AUTHORS	Carninci,P. and Hayashizaki,Y.				
AK008234	TITLE	High-efficiency full-length cDNA cloning				
AK008234	JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
AK008234	MEDLINE	99279253				
AK008234	POBMD	10349636				
AK008234	REFERENCE	2 (sites)				
AK008234	AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
AK008234	TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
AK008234	JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
AK008234	MEDLINE	2049374				
AK008234	PUBMED	11042159				
AK008234	REFERENCE	3 (sites)				
AK008234	AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwaigaki,K., Fujiwaka,S., Inoue,K., Togawa,M., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Yoneda,Y., Ishikawa,T., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.				
AK008234	TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer				
AK008234	JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				

QY 606 ctcccaaatgtctcattacagcaagtcggttggtgttactcctctcgtt-cca 664
|||||
Db 601 CTTCGCCAATGCTCTTCAAGCTAGCCACTAGCTGGTGGGTGTTATCTCTCTGTC 660
QY 665 ggttgcaagctgacctgaactgaactgaactgctccagagctca-gtgcggtgaa 723
|||||
Db 661 GGGTGCAGAGCTGACCTCAACTGACATGAGAGACTGTCTCCAGGTCCAGGCGGAA 720
QY 724 tactctcgtataacatgt-gggagcgactctcctataaagagcaatgcaga 773
|||||
Db 721 TACTTGTGAATTAATCATGTGGGAGCGGCTTCTATTAAGAGCATGGGAGA 771
RESULT 3
BG468641 731 bp mRNA linear EST 21-MAR-2001
LOCUS 602510243F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:464467 5',
DEFINITION mRNA sequence.
ACCESSION BG468641 GI:13400911
VERSION BG468641.1 GI:13400911
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 731)
NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM416 row: e column: 12
High quality sequence stop: 686.
location/Qualifiers
1. 731
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:464467"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 214 a 157 c 183 g 177 t
ORIGIN
Query Match 36.7%; Score 691.6; DB 10; Length 731;
Best Local Similarity 98.4%; Pred. No. 3.6e-190;
Matches 720; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

Db 122 AAGAGGCTATCACTTTAGGGGTACCCGAGGGGACAAAGGCAAGTTCACGGCT 181
QY 313 attctgaataacctggaggtcgaagaagcgaaagcttccacaagaccactactc 372
|||||
Db 182 ATTGTGAATTAACCTGGAGGCTCAAGAGAACGAGAGCTTTCACAGACACCTACTC 241
QY 373 tccctcacagagactgtgacacttcaagcgctgaagaagtcatatcagttccactg 432
|||||
Db 242 TCCCTCACAGAGACTGTGAGCACTTCAAGGCTGAAGAAAGTTCATACAGTCCACTG 301
QY 433 agcaaaagaagagtgaggttccctattgcatctctatgtgtattcatatgaagatgaa 492
|||||
Db 302 AGCAAAAGAGAGGTGGAGTTCCCTATTGCACTATGATGATGATGATGATGATGAA 361
QY 493 aacttgaagagctacgagcagctgtgtatgcccctcagaacatatactgtccatgtg 552
|||||
Db 362 AACTTTAAAGGCTACGACGAGCTGTATGCCCTCAGAAACATATCTGTGTCAGTGT 421
QY 553 gatgagaagtcgcccaagaacttccaaagagcggtcaaaagaattattctgtcccca 612
|||||
Db 422 GATGAGAGAGTCCCGAAGACTTCAAGAGAGCGGTCAAGCAATTAATTTCTCTCCCA 481
QY 613 aagttctcattagccagtaagctgtgtcgggtgttattcctcctcgtgtccaggtgcaa 672
|||||
Db 482 AATGCTCTCAATGACCACTAAGCTGTGGGTGGTTATCTCTGCTGTCAGGGTCAA 541
QY 673 gctgacctcaactgcaatgagaactgtctccagagctcagtgccgtgtgaaataactctg 732
|||||
Db 542 GCTGACTCAAGCTGACATGAGAACTTCTCCAGAGCTCACTGCGGTGAATATCTCTG 601
QY 733 aatacatgtggagcagacttctcctataaagaacatgacagatgtcagagctctcaag 752
|||||
Db 602 AATTCATGTTGGTGGAGCGGACTTCTATTAAGAGCATGCAGATGTGTCAGG--CTCAAG 659
QY 793 atgtgtgaatggaggaatagcatgaggt-cagaaggtacccctcagaacaaagaacacg 851
|||||
Db 660 ATGTTGAATGGAGCAATACATGAGGTACAGAGTACTCATTAAGCAAAAAGAAAACG 719
QY 852 ctggaatatca 863
|||||
Db 720 TTGGAATATTA 731
RESULT 4
BG470202 875 bp mRNA linear EST 21-MAR-2001
LOCUS 602533711F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:461355 5',
DEFINITION mRNA sequence.
ACCESSION BG470202 GI:13402477
VERSION BG470202.1 GI:13402477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 875)
NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LNCM460 row: e column: 04
High quality sequence stop: 814.
location/Qualifiers
1. 875
FEATURES
Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4661355"
/clone_lib="NH MGC.15"
/cisue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; site_1: XhoI; site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

```

	Query Match	35.4%	Score 666	DB 10	Length 875	
	Best Local Similarity	93.2%	Pred. No. 1.1e-182			
	Matches 766	Conservative 0	Mismatches 40	Indels 16	Gaps 6	
QY	1	gattgtgcctccctccacactccctcgtgtcgtgcctccacactgcgtcccatctctgacg	60			
DB	47	GATTGTGTCCTCCTCCACTTCCCTGTGCTGSGGTCTCCACTGTCTCCATTCTGTGAGC	106			
QY	61	atgtgtcaatggaagaagactctgcacgtgcataactctgtggctctggctgtctatctg	120			
DB	107	ATGTGTCAATGGAAGAGACTCTGCAGCGTCGATTACTGTGGGGCTCTGGGCTGCTATATG	166			
QY	121	ctgtcgtgcacactctgtgcctgaactctcttcacaggttgaaatgtacactctacacattg	180			
DB	167	CTGTGCGGCCTACTGTGCTCTGAACCTTTCTTTCAAGTTGAATGTGACTCTCAACCACTTG	226			
QY	181	ggtctggaagccaaagaaatctcaaaagccagtaactgttaagaaatacttgtataattccctg	240			
DB	227	GGTCTGAGATCCAGGGAATCTCAAGCCAGTACTGTAGGAATATCTGTATATTTCTCG	286			
QY	241	aaacttccaagcaagaaggtctcataactcgttcaaggtgtcaaccggaggagccaagaagca	300			
DB	287	AAACTTCCAGCAAAAGGCTATCAACGTTCAGGGGCTACCCGAGGGAGCCAAAGAGCA	346			
QY	301	gtgtctcagagctatctgaataacctgtgaggtcgaagaagaacggaagccttccacagac	360			
DB	347	GTGCTTCAAGCCTATTTCTGAATTAACCTGGAGAGCTCAAGAAAGACGAGAGCTTTCCACAGAC	406			
QY	361	accacactacactctccctccacagagactgtgacacttcaaaagctgaaagaaagttcata	420			
DB	407	ACCCACTACTCTCTCCCTCACAGAGACTGTGAGCACTTCAAGGCTGAAGGAAGTTCCATA	466			
QY	421	caattcccaactgagagaagaagaggtggaagttccctatctgcatctatctatgtgattcat	480			
DB	467	CAGTTCACACGTAGCAAAAGAGAGGTGGAGTTCCTATTATGCACTTAATGATGATTCAT	526			
QY	481	gagaagacttgaaaacttctgaaaggtctactcgaagctctgtatgtacccctcaagaacatatac	540			
DB	527	GAGAAGATTGAAGAACTTTGAAAGGCTACTGCGAGCTGTATATGCCCTCAGAACATATATAC	586			
QY	541	tgtgtccaatgtgatatgagaaggtccccaagaacatttcaaaagagcggtcaaaagcaattatt	600			
DB	587	TGTGTCCATGTGTGATGAGAAAGTCCCAAGAAACTTTCAAAAGAGCGGTCAAAACAATATATT	646			
QY	601	tctgtctcccaaatgcttctatagccaagtaagctgtgtctcgggtgtgttat---gcctcc	657			
DB	647	TCTTCTCTTCCCAAAATGTCTTCAATATCCAGTAAGCTGTGCTGGGTGGGTATATAGCCCTCT	706			
QY	658	tgtgtccagaggtgca---gctaacctcaaatctgcatgaa--gactgtccacagagcttccg	712			
DB	707	GGGTCCAGGAGTCAAGGCTTGACCTTCAACCTGCATAGGAAGAAAGACTTGCCCTCCAGAGGTGAG	766			
QY	713	tgcctgtggaata-cttccatgaatacatgttg-----acggacttccctataaagaaga	766			
DB	767	TGCCCTGGAAATACCTTCTCTGTAATACCTTGTGGGAGACCGAGACTTTTCTCATTTAAAGAGCA	826			

```
Qy      767  atgcag--agatgctcagcgtctcaagaatgttaatgag 806
        |||||  |||||  |||||  |||||  |||||  |||||
Db      827  ATGCAGCAGATTGGTCCACGGGTTCACAGATGTCGACACGGGG 868
```

RESULT	5
AA583339/c	
LOCUS	AA583339
DEFINITION	nm25f11.1 NC1_CGAP_GC5 Homo sapiens CDNA clone IMAGE:1086861 3' similar to TR:G176468 GI176468 ENZYMMATIC GLYCOSYLATION-REGULATING GENE 'r' mRNA sequence.
ACCESSION	AA583339
VERSION	AA583339.1
KEYWORDS	GI:2367948
SOURCE	EST.
	human.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 640)
AUTHORS	NCI-CCAG http://www.ncbi.nlm.nih.gov/ncicag .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: c3pb3s-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LMW at: www.bio.liml.gov/bdpp/image/image.html
Seq primer: ~40m3 fwd. ET from Amersham
High quality sequence stop: 454.
Location/Qualifiers
1..640
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:10886861"
/clone_1ib-"NCI_CGAP_GC5"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="vector: Bluescript SK-, Site_1: EcoRI; Site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dt. Mixed germ
cell tumors. 5' adaptor sequence: 5' GAATTCGCACAG 3' 3'
adaptor sequence: 5' CTCGACTTTTTTTTTTTTTTT 3' Average
insert size: 0.7 kb."

```

	Query Match	Similarity	34.0%	Score 640:	DB 9:	Length 640:
	Best Local	Similarity	100.0%	Prod. No.	3.7e-175:	
	Matches	640:	Conservative	0:	Mismatches	0:
				Indels	0:	Gaps
					0:	
QY	1189	gctccctatgctccctgctcgtcttggaaatccaaacaggggctatctcgcttatatgggctcgg	1248			
Db	640	GCCTCTTATGCTCCCTGCTGTGGAAATCCACGCGGGCTATCTGGTTATGGGCTGG	581			
QY	1249	gactctgaatttgatgcttcaaaacccatcacctgtgttgcgcaaaagtcttgatcccaaggta	1308			
Db	580	GACTGTGAATTGGATGCTCTCAAAACCATCATCCTGTTGGCGCAACAAGTTTGACCCAAGGTA	521			
QY	1309	gataataatgctcttcacatgctcttaagaagaataaccagcttataaggccatcatatagaact	1368			
Db	520	GATGATTAATGCTCTTCATGCTGCTTAAGAGATATCCTACCTTATTAAGCCATCATATGGACT	461			
QY	1369	gaacttttagacacacatltgaaagcgtttgtctactcgttggggcaagaagcatgtacaacat	1428			
Db	460	GAACTTTTGAGACACACTATGAGAGGCTTGCTACCTGCTGGGGCAAGACGATGTACAAACAT	401			

QY 1429 gctcagaactgctggagacagtgctggtaggagaccaggcttgcattcgtgacatc 1488
Db 400 GCTCAGAACTTGTCTGGAGAGTGTGGTGGAGACAGGGCTTTGCAATTCTGTGACATCC 341
QY 1489 tttagagtaagagggcgtcattatagattgctgggaagtagatcttggcccttgaattg 1548
Db 340 TTTAGATTAAGAGGGCGCTGCTATTGATTGTGGTAAGTAAGATCTTTTGGCTTGCATAATTG 281
QY 1549 ctgctgggtgtaatgctgcttctcctcaccaccaccaccaccaccaccaccaccacc 1608
Db 280 CTGCTGGGTGAATGCTGCTGTCTCTCAACCCCTAACCCTAGATGTTCCCTCCACTAAT 221
QY 1609 ttctcactatgtagaagtagaagtagatgtagaagtagaagtagaagtagaagtaga 1668
Db 220 TTCTCCTAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAG 161
QY 1669 agagcactgattcactgaatgaatgctgctgtagcttccaccctgtagagctgct 1728
Db 160 AGACCACTTGAATTCATTGAATGCCCTGCTGGTATGCTTTCCATTCCTGAGCTGCCCT 101
QY 1729 tccataaatccaggttggtagcgtggagaggaacttgaatggaagaagaccctccc 1788
Db 100 TCCATAAATATCCAGGTTGGTAGCCGTGAGAGAACTTGATGGAAGAAGAACTTCCC 41
QY 1789 ttctgactgttaactaaataaataagcttccgtatca 1828
Db 40 TTCTGTACTGTTAACCTTAATAAATAATAGCTCTGATTCA 1

RESULT 6

BE869192 842 bp mRNA linear EST 20-OCT-2000
LOCUS 601445191F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3849215 5',
DEFINITION mRNA sequence.

ACCESSION BE869192
VERSION BE869192
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 842)
AUTHORS NIH-MGC
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LLAM9566 row: e column: 24
High quality sequence stop: 684.
Location/Qualifiers

FEATURES
source 1. 842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3849215"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies.

BASE COUNT

234 a 190 c 209 g 209 t

ORIGIN

Query Match

33.6%; Score 633.4; DB 10; Length 842;

Best Local Similarity 93.8%; Pred. No. 3.6e-173;
Matches 736; Conservative 0; Mismatches 41; Indels 8; Gaps 7;

QY 442 gaagtgaagtcctcattatgcatatctatgtagtattcatgagaagattgaaacttga 501
Db 1 GAGGTGAGAGTTCCTTA-TGCATACCTCATGTGATTATGAGAGA-TGAATACTTTGAA 58
QY 502 aggtactgagagctgtagtgcctccctcagaacataactgtagtgcattgtagaag 561
Db 59 AGGCTACTGCGAGCTGTGTATGCCCCCAGAAATATATACCTGTCCATCTGTGATGAGAA 118
QY 562 tccccagaacttccaaagagcggtcaagcaatattctgtcttcccaatgcttc 621
Db 119 TCCCGAAGAACTTCAAGAGCGGTCAAGCAATTAATTTCTTCTTCCCAATGCTTC 178
QY 622 atagccagtaagctgtagtgcctcctcctcctcctcctcctcctcctcctcctc 681
Db 179 ATGACGAGTAAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 238
QY 682 aactgcatggaagactgctgcagagctcagtgctgtagaataactcctgaatcagt 741
Db 239 AACTGCATGGAAGACTTGTCTCCAGAGCTCAGTCCCGGAAATACCTTCTGAAATACATGT 298
QY 742 ggagcagacttctcctataaagcaatgcaagatgtagcaggctcctcaagatgtagat 801
Db 299 GGGAGGAGCTTCTCTATTAAGAGCAATGCAAGATGCTCCAGGCTCTCAAGATGTTGAAT 358
QY 802 ggagagatagcatgtagtgcagagtagcctcctcctcctcctcctcctcctcctcctc 861
Db 359 GGGAGGAGATGATGAGAGTGCAGAGTACCTCTCAAGACAAAGAACCCCTGGAATAT 418
QY 862 caacttgagtagttagagacacattacacctaaccacaagaagaagatcctccct 921
Db 419 CACTTGTAGGTAGTGAAGACACATTAACCTTAACCAAGAAAGATCTCTCCCTT 478
QY 922 tataatttaactatggttaaggaagcgtagaattgtagctccgaagattcgtccaa 981
Db 479 TATTAATTAATGATGTTAAGGGAATGCGTACATGTTGCTTCCGAGATTTGCTCAA 538
QY 982 catg-ttttgaagaacccataaccacaactgattgaatggtgtaaaagacac-ttata 1039
Db 539 CAGTGTTTTGAAGAACCCCTTAATCCCAACAGATTTGATGATGATGATGATGATGATGAT 598
QY 1040 gcccaagatgaacacctctggtgcaacccctcagctgtagcaggtgtagtgcctgctgttc 1099
Db 599 GCCCAGATGAAGAACCTCTGTGGCCACCTTCAACCGTGCAGCGTGAATCCCTG-TCGTTC 657
QY 1100 ccaaccaccccaagtagacatctcagacatgacatctcattgccaagctgtagaagtg 1159
Db 658 CCAAC-ACCCAGATGACGATCTCAGACATGATTTCTATTGCGCGCTGATTAATGGGC 715
QY 1160 aggtcatgagggagacatcgatagagtgtagtgcctcctcctcctcctcctcctc 1219
Db 716 CGGTCAATGAGGAGCAATTCGATTAAGGCGTGTATATGCTCTTGT-TCGGAATCCAC 774
QY 1220 aacggg 1224
Db 775 ACGGG 779

RESULT 7

AW842622/c 655 bp mRNA linear EST 18-MAY-2000
LOCUS AW842622
DEFINITION MR2-CN0035-220300-204-f09 CN0035 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW842622
VERSION AW842622.1 GI:7936605
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 655)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

TITLE
JOURNAL
MEDLINE
COMMENT

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-MR2-CN0035-220300-204-f09&f3=2000-03-22&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 652.

FEATURES

source

```
1. .655
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0035"
/dev_stage="Adult"
/note="Organ: colon, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
```

BASE COUNT

151 a 162 c 152 g 190 t

ORIGIN

Query Match 33.0%; Score 622.6; DB 9; Length 655;
Best Local Similarity 98.6%; Pred. No. 4.4e-170;
Matches 628; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

Oy 275 gggtcaccgaggagcccaagaagcagtgctcagcctctcgaataacttgagagca 334
    |||||||
Db 637 GGGTCACCCGAGGGGACCAAGAGGCGAGTCTTCAGGCTATTCTGAATACCTGGAGGTCA 578

Oy 335 agaagaagcgaagccttcacagacacccactccctccacagagacttgagc 394
    |||||||
Db 577 AGAGAAACGAGAGCGCTTTCACAGACACCACTACCTCTCCCTCACGAGACTGTGAGC 518

Oy 395 actccaagctgaagaagagtcatacagttcccaactgagcaagaagagtgagttcc 454
    |||||||
Db 517 ACTTCAGAGCTGAAGAGAGTTCATACAGTCCACCTGAGCAAGAAAGAGTGAGTTCC 458

Oy 455 ctattgcatactatagtgatcataggaagattgaaacttgaaagagctactcgag 514
    |||||||
Db 457 CTATTGCACTCTATGTTGTTGATTCATGAGAGATTGGAAGCTTTGAAAGGCTACTGCGAG 398

Oy 515 ctgtgtatgcccccaagaacatactatgcatgtgcatgtgagagaagccccagaactt 574
    |||||||
Db 397 CTGTGTATGCCCTCAGAGACATATACTGTGTCCATGTGATGAGAGAGGCCCAAGAACTT 338

Oy 575 tcaagaaggggccaagaacatttcttgctccccaatgctccatagccagtaagc 634
    |||||||
Db 337 TCAAGAAGAGCGGCTCAGCAATTAATTTCTTCTCTCCCAATGCTCTCATAGCCAGTAGAC 278

Oy 635 tggttcgagtggttatagccccctggtccagagtgcaagcgcgaactcaactgagtagag 694
    |||||||
Db 277 TGGTTCCGGGTGTTATGCCCTGCTGTCAGGAGGTCAGAGCTGACCTCAACTGATGAGAG 218
```

```

Oy 695 actgtccagaagctcagtcgctggaataactctctgaatacatgttgagcaggacttc 754
    |||||||
Db 217 ACTGCTCCAGAGCGTCATGCGCGTGAATATCTTCTTAATACATGTGGAGCGACTTTC 158

Oy 755 ctataaagacgaatgcaagaatggtcagcgtctcgaagtgttgagtgaggaatagca 814
    |||||||
Db 157 CTATTAAGAGCAATGCAAGATGTGTCAGGCTCTCAAGATGTGTAATGGAGGTATACCA 98

Oy 815 tggagtcagagtgctcctctcctaagacaaagaaacccgttggaatatcattgagtag 874
    |||||||
Db 97 TGAAGTCAGAGTACTCTCTTAAGCACAAGAAACCCGCTGGAATATCATTGTGAGGTAG 38

Oy 875 tgagagacacattcacctaaccaacaaagaagaagga 911
    |||||||
Db 37 TGAGAGACACTTACACTTACCAACAATATTCACCA 1
```

RESULT 8
BG469448 724 bp mRNA linear EST 21-MAR-2001
LOCUS 602532948F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4660835 5',
DEFINITION mRNA sequence.
ACCESSION BG469448
VERSION BG469448.1 GI:13401723
KEYWORDS EST.
SOURCE human.
ORGANISM human.

REFERENCE 1 (bases 1 to 724)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: NIH Intramural Sequencing Center
clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LNCM458 row: O column: 12
High quality sequence stop: 687.

FEATURES

source

```
1. .724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4660835"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
```

BASE COUNT

180 a 176 c 179 g 188 t

ORIGIN

Query Match 32.5%; Score 613.2; DB 10; Length 724;
Best Local Similarity 95.9%; Pred. No. 2.5e-167;
Matches 694; Conservative 0; Mismatches 23; Indels 7; Gaps 6;

```

Oy 7 gtctctccagccttccctgtg-ctcggtccacccggtcccatctctgagagatggt 65
    |||||||
Db 2 GTCTCTCTCCACCTTCCTCTGTCCTGCTGTCACCTGTCTCCATTTCTGTGAGAGATGT 61

Oy 66 tcaatggaagagactctgcagcgtcatcttgtaggctctggtgctatcatgctgct 125
```

```

|||||
Db 62 TCATGGAAGAGACTGTGACGTGATTAATCTGTGGCTGTATATGCTGCT 121
|||
Qy 126 ggcacatgtgctctgaacattctcttcaggttgaaagtgcacacactgggtc 185
|||
Db 122 GGGCACTGTGGCTGTAACATTCTTTCAAGTGTGAAGTGTGACTGTGACACTGGGTCT 181
|||
Qy 186 ggaagtcagggaactctcaagccagtaactgttagaatactctgtataattccatgaact 245
|||
Db 182 GGAGTCCA-GGAATCTCAAGAGCCAGTACGTAGAAATATCTTATATATTTCTGAAACT 240
|||
Qy 246 tccaagaaagaggtctatcaactgttcagggtgcacccgaaggagaccagaagcaatgct 305
|||
Db 241 TCCAGGAAGAGGCTTATCAACTGTTCAAGGGGTCACCCGAGGGACCAAGAGGACAGTGT 300
|||
Qy 306 tcaggtatcttaaacactcgtgaaggtcaagaagaacgagaccttcacaacacccc 365
|||
Db 301 TCAGGCTATTCTGATTAATTAACCTGAGAGTCAAGAAAGCCAGAGCCTTTACAGACACCA 360
|||
Qy 366 ctactctccctcaccagagactgtgagcaactcaaggtcgaaggaggtcatalactat 425
|||
Db 361 CTACTCTCCCTCCACAGAGACTGTGACACTTCAAGGCTGAAGAGAAAGTTATACAGTT 420
|||
Qy 426 ccactcagcgaagaagaggttgaggtccctatgtcactctctatgtgtatagaag 485
|||
Db 421 CCACGTGAGCAAGAGAGGAGGTGAGTCCCTATTTGATATCACTATGCTGATGATAGAA 480
|||
Qy 486 gattgaaacttgaaagcctactgcagctgtgtatgccccacagaa-catalactgtg 544
|||
Db 481 GATTGAAACTTTGAAAGGCTACTGCGAGCTGTGTATGCCCTCACAACCATATACTGTG 540
|||
Qy 545 tccatgtgatgaggaagtcgccca--gaaacttcaagaagggcgccaagaactatctc 602
|||
Db 541 TCCATGTGATGAGAGTCCCAAGAAAGATTTTCAAGAGGCGGTG-AAAGCAATTTATTC 599
|||
Qy 603 ttgctcccaaatgtcttaataagccag-taagctgtgtcgggtgtgttcctcctgtgt 661
|||
Db 600 TTGCTTCCCAAAATGTCTTATACCCAGTTAAGCTGTGCGGTGTGACATGCTCCGTGG 659
|||
Qy 662 ccaggtgtcgaagctgacactcaactgcatgtgaagactgccacacacacagtcagtcg 721
|||
Db 660 TCCAGGTGCAAGCTGACCAAAATGATGGAGCATTGGCTACAGACTCAGTCCGTGGA 719
|||
Qy 722 aata 725
|||
Db 720 AATA 723
|||

RESULT 9
LOCUS BE292814 612 bp mRNA linear EST 13-JUL-2000
DEFINITION 601105179P1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:2988090 5',
ACCESSION BE292814
VERSION BE292814.1 GI:9175472
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 612)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bcr-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM76 row: m column: 19

```

```

FEATURES
source High quality sequence stop: 605.
location/Qualifiers
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2988090"
/clone_1ib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB7; Site: 1: XhoI; Site-2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 181 a 137 c 140 g 154 t
ORIGIN
Query Match 32.4%; Score 611; DB 10; Length 612;
Best Local Similarity 100.0%; Pred. No. 1e-166;
Matches 611; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 455 ctattgcatactctatggtgatcatgagaagattgaaacttgaaagctactgcag 514
|||
Db 1 CTTATTCATATCTATGATGATTCATGAGAAATGAAAACCTTGAAAGGCTACTCGAG 60
|||
Qy 515 ctgtgtatgccccctcagaacatactgtgtcgaatgtgagtagaagatccacagaact 574
|||
Db 61 CTGTGTATGCCCTCGAAGCATATATGTTGCTCATATGTGATGAGAGTCCCAAGAAACT 120
|||
Qy 575 tcaagaagcggtcaagaacatattctctgtctcccaaatgtctcatagccaagaagc 634
|||
Db 121 TCAGAAAGGGGGTCAAGACATATTTCTTGTCTCCCAATATGCTCATAGCCAGTAAGC 180
|||
Qy 635 tgggtcgggtgtgttatgctctccctcgtgccaggggtcgaagctgaactcaactgata 694
|||
Db 181 TGGTTCGGGTGTTTATGCTCTCCTGTCAGAGGTGCACCTGCATCATGATGAGAG 240
|||
Qy 695 actgtctcagaagctgaagtcgggtgaaatctctctgaatacatgttggaaggaacttc 754
|||
Db 241 ACTTGCTCCAGAGCTCAGTGCCGTGGAATATCTTCTGTAATCATGTGGAGCGACTTTC 300
|||
Qy 755 ctataagaagcaatgcagatagtgctcaggctctccaaagtgtgaaatgggaagaataga 814
|||
Db 301 CTATAAAGAGCAATGCAGAGATGTGTCAGGCTCTCAAGATGTTGAATGGAGGAATAGCA 360
|||
Qy 815 tggagtcagaggtactctcctaaagacaagaagaacccgtctggaataatcaacttgagt 874
|||
Db 361 TGGAGTCAGAGGTACTCTCTTAAGCACAAAGAAACCCGCTGGAAATATCACTTGGAGT 420
|||
Qy 875 tgaagacacattacacccaaacaaagaagaagatctccccccttaattatattaca 934
|||
Db 421 TGAGAGACACATTACCTTACCTCAACAGAAAGAGATCTCCCTTTAATTTAATTAAC 480
|||
Qy 935 tgtttacaggaatgcgtacatgtgtgtctccgaattctgtccacatgttttgaaga 994
|||
Db 481 TGTTTACAGGAATGCGTACATTTGTGCTCCGAAATTTGCGCCAACTGTTTGAAGA 540
|||
Qy 995 acccttaatccacaacatgattgaatgggtlaaagaacattatagccagatgaaacacc 1054
|||
Db 541 ACCCTTAATCCCAACATGATGATGGGTAAAGACACTTATAGCCAGATGAAACACC 600
|||
Qy 1055 tctggggcacc 1065
|||
Db 601 TCTGGGCCACC 611
|||

RESULT 10
LOCUS BE867668 1050 bp mRNA linear EST 20-OCT-2000
BE867668

```

DEFINITION	60144320661 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847441 5', mRNA sequence.
ACCESSION	BE667668
VERSION	BE667668.1 GI:10316444
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 1050)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L1AM9561 row: 1 column: 02 High quality sequence stop: 599. Location/Qualifiers 1. 1050 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3847441" /clone_1id="NIH_MGC_65" /tissue_type="adenocarcinoma" /lab_host="DH10B (phage-resistant)" /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NciI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."
BASE COUNT	390 a 242 c 231 g 187 t
ORIGIN	
Query Match	32.2%; Score 607.2; DB 10; Length 1050;
Best Local Similarity	86.9%; Pred. No. 1.8e-165;
Matches 702; Conservative	0; Mismatches 103; Indels 3; Gaps 3
OY	357 agacaccactacccctccctcaccagagactgtgagcactcgaagcgtgaagaagat 416
Db	1 AGACACCCACACTCCTCCCTCACAGAGACNCTGTGACACTTCAAGGCTGAAGAAGATT 60
OY	417 catcaagttccacacgagagaagaagagtgtaagttccctattgtactatagtgtat 476
Db	61 CATACAGTTCCACATGAGCAAGAAAGAGGTGAGTTCCTATTGCATCTATAGTGTGAT 120
OY	477 tcatagagaagattgaaaacttcgaaagcgtactgcagctgtgtatgccctcagaacat 536
Db	121 TCATGAGAGAAGATTGAAAACCTTTGAAAAGGCTACTGCAGCTGTGTATGCCCTCAGAACAT 180
OY	537 atactgtgtcactgtgtagtagaagaatgcccaaaaactttcaagaagcgcgtcaaaagcaat 596
Db	181 ATACTGTGTCCATGTGAGTAGAGAGAGCCCAAAACTTTCAAAAGAGCGGTCAAAACAT 240
OY	597 tattcttgcgtcccaaatgatcttcatagcaagtaagctggtgtcgggtgtttagtcctc 656
Db	241 TATTTCCTGGCTTCCCAAAAGCTTCTCATACCCAGTAAGCTGAGTGGGTGGTATTATCCCTC 300
OY	657 ctgtgccaaggtgtcaagctgacactcaactgacatgtagaagactgtgtccagagctcagttgcc 716
Db	301 CTGTGTCAGAGGTGCAAGCTGACCTCAACTGCATGGANACTTGCTCCAGAGCTCAAGTGGC 360
OY	717 gtggaataacttctcgtaatatcatgtgtgagcggagcttccctataaagagcaatgtcaagat 776
Db	361 GTGGAATATCTTCTCTGAATACATGTGTGGAGCGGACTTTCTATPAAGCAGCAATGCAAGAT 420
OY	777 ggtccagagctccaagatgttgaatgaggagaatagcatgtgaagtgcaaggttacctcttaa 836

[illegible]

Query Match 32.1%; Score 605; DB 9; Length 671;
 Best Local Similarity 97.1%; Pred. No. 6e-165;
 Matches 647; Conservative 0; Mismatches 16; Indels 3; Gaps 3;

QY 1168 gagggagacatgataaggtgtctcttattgtctcctgtctctgtctgtgaatccacgagggct 1227
 |||||
 Db 671 GAGGAAGACACGATTAAGGGGTGCTCTTAAGCTCCCTGCTCTGGAAT-CACCAAGGGGGCT 613
 QY 1228 atctggttatatgggctgagggactggaattgtatgtcttcaaacaccacccgttggcc 1287
 |||||
 Db 612 ATCTGGCTTTATGCGGCTGGGGACTTGAATGATGTC-TCAAAACATCACCTCTTGCC 554
 QY 1288 aacaagttgaccccaagtagatgaatgtaatgctcttcagtgtcttagaagaataacctagt 1347
 |||||
 Db 553 AACCAATTGACCCCAAGAGAGATGATGTAATGCTCTCAGTGGTTAGAGAAATACNT-CGT 495
 QY 1348 tataaagccatctatgtggaatgaatttgaagaacacatagagaaggtgtctctgttg 1407
 |||||
 Db 494 TATTAAGCCATCTATGAGGACTGCTTTGAGACACTATGAGAGCGTTGCTACTGTG 435
 QY 1408 ggtcaagacatgacacaaatgctcagaacttgcagacagtgtgtgtgagagaccag 1467
 |||||
 Db 434 GCGAAGAGCATGTACAACATGCTCAGAACTTCTGGAGAGTGTGGGTGGAGACCAG 375
 QY 1468 gcttgcgaatctgtgacatctttagataagaagggctgctatagattgtgtgtaagta 1527
 |||||
 Db 374 GCTTGCATATCGTGCAATCCTTGAATGAGAGGCGCTGCTATTAGATGTGGGTAAATA 315
 QY 1528 gatcttgcctgctgaattgtcgtcgtggtgtaagtctgttcttcttccaccctaac 1587
 |||||
 Db 314 GATCTTTTGGCTTGCAAAATGTGCTGCGGTGATGCTGCTTCTCACCCTTAACC 255
 QY 1588 ctatgagtcctccactaacttctcactaagtagagaatagaaactgtctgtatagagag 1647
 |||||
 Db 254 CTAGTAGTTCCTCCTACTAATCTTCTCATAAGAGATAGAACTGCTGTGATAGGAG 195
 QY 1648 agtgaaggaaggaatgtagtagaagaacttgattcaagttgaatgctgtgtagctt 1707
 |||||
 Db 194 AGTGAAGAGAGGATGATGCTAGACACTGATTTGATTTGATTTGATTTGATTTGATTTG 135
 QY 1708 tcatctctgtgagagctgctcgtctccttaataatctcaggttgtgtaggtgagagaact 1767
 |||||
 Db 134 TCCATTTCTGTGAGCTGCCCTTCTTAATTAATTCAGGTTGGTAGCGTGAGAGAACTT 75
 QY 1768 tgaaggaaagagaaccttccctctgtactgttaacttaaaaaataaagctcctgattc 1827
 |||||
 Db 74 TGATGAAGAGAACTTCCCTCTGTAAGTAACTTAATAATAATTAATTAATTAATTAATTA 15
 QY 1828 aaagta 1833
 |||||
 Db 14 AAAGTA 9

RESULT 12
 B1765410 766 bp mRNA linear EST 25-SEP-2001
 LOCUS B1765410 603050445F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5190611 5',
 DEFINITION mRNA sequence.
 ACCESSION B1765410
 VERSION B1765410.1 GI:15756988
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 766)
 NIH-MGC http://mhc.mcg.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@b-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

CDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1476 row: 1 column: 12
 High quality sequence stop: 766.
 Location/Qualifiers
 1..766
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5190611"
 /clone_11b="NIH_MGC_116"
 /lab_host="DH10B"
 /note="Organ: pooled colon, kidney, stomach; Vector: pCMV-Sport6; Site:1; Nott; Site:2; EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon, 46 yo female kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."

BASE COUNT 199 a 184 c 192 g 190 t 1 others
 ORIGIN

Query Match 32.0%; Score 602.6; DB 10; Length 766;
 Best Local Similarity 98.9%; Pred. No. 3.2e-164;
 Matches 638; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 1 gatgtctcctccacaccttccctgtcgtgctcagctgtccacgtgtccatctgtgag 60
 |||||
 Db 120 GAGTGTCTCTCTCCACCTTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 179
 QY 61 atgttcaatlgagaagactctgcagctgacattctgtgtgtgtgtgtgtgtgtgtgtgtgt 120
 |||||
 Db 180 ATGTTCAATGGAAGAGACTCTCCAGCTGATTAATTTGGGCTGTGGCTGTGCTGTATAG 239
 QY 121 ctgtctgcaactgt 180
 |||||
 Db 240 CTGCTGGCAGCTGTGCTCTGAAACTTTCTTCAAGTTGAAGTGTGACCTGTGACCACTTG 299
 QY 181 ggtctgagatccaggaatctcaagccagtaactgaagaaatctctgataatctctgt 240
 |||||
 Db 300 GGTCTGAGTCCAGGGAATCTCAAGCCAGTACTGTAGGAAATTTCTTATATTTCCTG 359
 QY 241 aaactccagcaaaaggtctatcaactgttcaagggtcaaccgagggagcaagaagca 300
 |||||
 Db 360 AAACCTCCAGCAAAAGAGGCTTATCAACTGTTCAGGGGTACCCGAGGAGCAACAAAGAGCA 419
 QY 301 gttgtcaggtctatttgaataacctgaggtgtcaagaaggaagagcttccagagc 360
 |||||
 Db 420 GTGCTTCAGGCTATTCTGATAACCTGTGAGGTCAAAAGAGGAGAGCTTTCAACAGAC 479
 QY 361 acccaactaccttccctccacagagactgtgaagactcaaggtcaaggtcaaggtcaag 420
 |||||
 Db 480 ACCCACTACTCTCCCTTCACACAGAGACTGTGAGCACTTCAAGGCTCAAAAGAGAGTCA 539
 QY 421 cagttcccaactgagcaaaagaggtgaggttccctattgcatctatgtagatcat 480
 |||||
 Db 540 CAGTTCCTCCACTGAGCAAAAGAGGAGGTGAGTTCCTATTCATCTATGAGTTCAT 599
 QY 481 gagaagatgaaaacttgaagagctactgagagctgtgtatagccctca-gaactata 539
 |||||
 Db 600 GAGAAATGAAACTTTAAAGGCTACGCGAGTGTATGAGCCCTCAGGAACATATA 659
 QY 540 cttt-gtccatgtgagtaggaagttcccaagaaacttcaag-aggcgggtcaagaagatt 597
 |||||
 Db 660 CTGTGTCTATGTGATGAGAGTCCCAAGACACTTCAAAAGAGGCGGTCAAGCAATT 719

QY 598 attcttgcctcccaaatgtcctcatagccaagtacgtylltcgg 642
|||||
Db 720 ATTCTTGCTTCCCAATGTCTTCATAGCCACGTAAGCTGGTCGGG 764
|||||

RESULT	LOCUS	DEFINITION
13 AM192901/c	603 bp	linear EST 29-NOV-1999
LOCUS	AM192901	x155b1.2.x1 NCI-CGAP Pan1 Homo sapiens CDNA clone IMAGE:4678591 3'
DEFINITION	similar to TR:O95395 O95395.BETA.1, 6-N-ACETYLGALUCOSAMINYLTRANSFERASE.	; mRNA sequence.

BASE COUNT	176 a	158 c	119 g	150 t
ORIGIN				

Query Match	31.8%	Score 559.8	DB 9	Length 603
Best Local Similarity	99.7%	Pred. No. 1.9e-163		
Matches 601; Conservative	0	Mismatches 2	Indels 0	Gaps 0

QY. 1285 gccacaagttgacccaagtagatgataatgtctcttcagtgcttagaagaataccta 1344

1345 cgtataagccatctatggaacttggagacacactatgagagcgttgtaactg 1404

[illegible]

1465 agggccttgcgaattcgtgycatcccttaagataagaaggctgcctatagattgtggrtaa 152
 153 |||||
 154 |||||
 155 |||||
 156 |||||
 157 |||||
 158 |||||
 159 |||||
 160 |||||
 161 |||||
 162 |||||
 163 |||||
 164 |||||
 165 |||||
 166 |||||
 167 |||||
 168 |||||
 169 |||||
 170 |||||
 171 |||||
 172 |||||
 173 |||||
 174 |||||
 175 |||||
 176 |||||
 177 |||||
 178 |||||
 179 |||||
 180 |||||
 181 |||||
 182 |||||
 183 |||||
 184 |||||
 185 |||||
 186 |||||
 187 |||||
 188 |||||
 189 |||||
 190 |||||
 191 |||||
 192 |||||
 193 |||||
 194 |||||
 195 |||||
 196 |||||
 197 |||||
 198 |||||
 199 |||||
 200 |||||
 201 |||||
 202 |||||
 203 |||||
 204 |||||
 205 |||||
 206 |||||
 207 |||||
 208 |||||
 209 |||||
 210 |||||
 211 |||||
 212 |||||
 213 |||||
 214 |||||
 215 |||||
 216 |||||
 217 |||||
 218 |||||
 219 |||||
 220 |||||
 221 |||||
 222 |||||
 223 |||||
 224 |||||
 225 |||||
 226 |||||
 227 |||||
 228 |||||
 229 |||||
 230 |||||
 231 |||||
 232 |||||
 233 |||||
 234 |||||
 235 |||||
 236 |||||
 237 |||||
 238 |||||
 239 |||||
 240 |||||
 241 |||||
 242 |||||
 243 |||||
 244 |||||
 245 |||||
 246 |||||
 247 |||||
 248 |||||
 249 |||||
 250 |||||
 251 |||||
 252 |||||
 253 |||||
 254 |||||
 255 |||||
 256 |||||
 257 |||||
 258 |||||
 259 |||||
 260 |||||
 261 |||||
 262 |||||
 263 |||||
 264 |||||
 265 |||||
 266 |||||
 267 |||||
 268 |||||
 269 |||||
 270 |||||
 271 |||||
 272 |||||
 273 |||||
 274 |||||
 275 |||||
 276 |||||
 277 |||||
 278 |||||
 279 |||||
 280 |||||
 281 |||||
 282 |||||
 283 |||||
 284 |||||
 285 |||||
 286 |||||
 287 |||||
 288 |||||
 289 |||||
 290 |||||
 291 |||||
 292 |||||
 293 |||||
 294 |||||
 295 |||||
 296 |||||
 297 |||||
 298 |||||
 299 |||||
 300 |||||
 301 |||||
 302 |||||
 303 |||||
 304 |||||
 305 |||||
 306 |||||
 307 |||||
 308 |||||
 309 |||||
 310 |||||
 311 |||||
 312 |||||
 313 |||||
 314 |||||
 315 |||||
 316 |||||
 317 |||||
 318 |||||
 319 |||||
 320 |||||
 321 |||||
 322 |||||
 323 |||||
 324 |||||
 325 |||||
 326 |||||
 327 |||||
 328 |||||
 329 |||||
 330 |||||
 331 |||||
 332 |||||
 333 |||||
 334 |||||
 335 |||||
 336 |||||
 337 |||||
 338 |||||
 339 |||||
 340 |||||
 341 |||||
 342 |||||
 343 |||||
 344 |||||
 345 |||||
 346 |||||
 347 |||||
 348 |||||
 349 |||||
 350 |||||
 351 |||||
 352 |||||
 353 |||||
 354 |||||
 355 |||||
 356 |||||
 357 |||||
 358 |||||
 359 |||||
 360 |||||
 361 |||||
 362 |||||
 363 |||||
 364 |||||
 365 |||||
 366 |||||
 367 |||||
 368 |||||
 369 |||||
 370 |||||
 371 |||||
 372 |||||
 373 |||||
 374 |||||
 375 |||||
 376 |||||
 377 |||||
 378 |||||
 379 |||||
 380 |||||
 381 |||||
 382 |||||
 383 |||||
 384 |||||
 385 |||||
 386 |||||
 387 |||||
 388 |||||
 389 |||||
 390 |||||
 391 |||||
 392 |||||
 393 |||||
 394 |||||
 395 |||||
 396 |||||
 397 |||||
 398 |||||
 399 |||||
 400 |||||
 401 |||||
 402 |||||
 403 |||||
 404 |||||
 405 |||||
 406 |||||
 407 |||||
 408 |||||
 409 |||||
 410 |||||
 411 |||||
 412 |||||
 413 |||||
 414 |||||
 415 |||||
 416 |||||
 417 |||||
 418 |||||
 419 |||||
 420 |||||
 421 |||||
 422 |||||
 423 |||||
 424 |||||
 425 |||||
 426 |||||
 427 |||||
 428 |||||
 429 |||||
 430 |||||
 431 |||||
 432 |||||
 433 |||||
 434 |||||
 435 |||||
 436 |||||
 437 |||||
 438 |||||
 439 |||||
 440 |||||
 441 |||||
 442 |||||
 443 |||||
 444 |||||
 445 |||||
 446 |||||
 447 |||||
 448 |||||
 449 |||||
 450 |||||
 451 |||||
 452 |||||
 453 |||||
 454 |||||
 455 |||||
 456 |||||
 457 |||||
 458 |||||
 459 |||||
 460 |||||
 461 |||||
 462 |||||
 463 |||||
 464 |||||
 465 |||||
 466 |||||
 467 |||||
 468 |||||
 469 |||||
 470 |||||
 471 |||||
 472 |||||
 473 |||||
 474 |||||
 475 |||||
 476 |||||
 477 |||||
 478 |||||
 479 |||||
 480 |||||
 481 |||||
 482 |||||
 483 |||||
 484 |||||
 485 |||||
 486 |||||
 487 |||||
 488 |||||
 489 |||||
 490 ||

QY	1525	gtaaatccttgccttggaatctgctgcctggatgagctgtgtcttcaacccta	1594
Db	303	GTAAATCTTTGGCCCTGGAAATGCTGCTGGGGAAAGCTGCTGTCTCAACCCCTA	244
QY	1585	accctagtagtctccctcaactcttcctcaatgtgaatggaatggaactgtctgataag	1644
Db	243	ACCCTAGTAGTTCCTCCCACTTAATTTCTCATAGTGAATGGAATGGAATGCTGTGATAGG	184
QY	1645	gagagtgaaaggagagatattgtgtagaagcaactgattcagttgaaatgctcgtcgtgtagc	1704
Db	183	GAGAGTGAAAGGAGGATATGTGTGTAGACCACTTGATTTCACTGAATGACCTGTGTGTAGC	124
QY	1705	ttttccattctgtgtagcgtcgcgtttcctcaataatccaagtttggtagcgtggaagaa	1764
Db	123	TTTTCCATTCTGTGTGAGCTGCCGTTCCAAATAATTCACAGCTTGTGTGTGAGCGAGGAGAA	64
QY	1765	ctttgatggaaaggaaacctcccttcctgtactgttcaactaaataaataagctccga	1824
Db	63	CTTTGATGGAAAGAGAACCTTCCCTTGTTACTGTTAACTTAAAAATCAATAGATCCCTGA	4
QY	1825	ttc	
Db	3	TTC	1

RESULT	1A
AM662377/c	
LOCUS	606 bp mRNA linear EST 06-APR-2000
DEFINITION	h125h05.x1 NCI CGAP CO14 Homo sapiens CDNA clone IMAGE:2973369 3'
	similar to TR:O95335 O95335 BETA1 .6'-N-ACETYLTGUCOSAMINYLTRANSFERASE.; mRNA sequence.

BASE COUNT	177 a	157 c	121 g	150 t	1 others
ORIGIN					
Query Match		30.98;	Score 582.6;	DB 9;	Length 606;
Best Local Similarity		97.5%;	Pred. No. 1.9e-158;		

Matches	591;	Conservative	0;	Mismatches	15;	Indels	0;	Gaps	0;
QY	1222	cgggcatcctgcttataaggagctgggagcttgaaatgagcttcaaaaccacccctg	1281						
Db	606	CGGCTATCCGTTATTATGAGCTGGGAGCTTCATTGATGCTTCAACACCATCACTG	547						
QY	1282	ttggccaacagcttgaacccaagaagtagatgataatgctcttcaagtcttaagaagaatc	1341						
Db	546	TTGGCTAACAAAGTTGACCCATGAGTAGAATGAATGCTTCTAGTCTTGAAGAATAC	487						
QY	1342	ctacgtataaaggccatctatgagagcttgagacacttgagagagcggtctctac	1401						
Db	486	NTACGCTTATAGGCCATCTATGGAATGAGCTTGAGACACACTATGAGAGCGCTGTAC	427						
QY	1402	ctgttgaggcaagagcatgtcaacaactgtcagaacttgcctggagacagtgaggagag	1461						
Db	426	CTGTGGGCGCAAGAGCATGTATCAAAACATGCTCAAGACCTGCTGGAGACAGTGGTGAG	367						
QY	1462	accagggcttgcgaatcgtgacatcccttaagataagagggctgctctatagattgagg	1521						
Db	366	ACCAGGCTTTGCAATTCGCGCATCCTTATAGATTAAGAGGCGTGTATAGATTGGG	307						
QY	1522	taagtagatcttgccttgcaaatgtgcctgggtgaatgctgtctgtctcacc	1581						
Db	306	TAACTAGATCTTTGCTTGCCTTCAAAATGCTGCGGTGAATGCTGCTTCTCTCACCC	247						
QY	1582	ctaacctagtagtctcctcactaacttctcactaagtgaagtagaactgctgtat	1641						
Db	246	CTAACCTAGTAGTCTCTCCACATCTTCTCTCACTAAGTAGAATGAGACCTGCTGAT	187						
QY	1642	agggaagaggaaggaaggaatgtagtagagcacttgattcagtgaaagcctgctgct	1701						
Db	186	AGGGAAGAGTGAAGAGAGGATGTGTAGAGCATGATTGATTCAGTTGAATGCTGCTGAT	127						
QY	1702	agcttccatctctgtggagctgcctcccttaataatccaggttggtagctggagga	1761						
Db	126	AGCTTTTCATCTCTGTGAGAGCTGCCCTTCTATTAATTCAGGTTTGTGTGCGTGAAGA	67						
QY	1762	gaactttagatgaagaagaacctccctctctgtactgttaacttaaaataatagctc	1821						
Db	66	GAACTTGTAGTGAAGAAGAACCTTCCCTTCTGTACTGTTAACCTAAATAATAGCTCC	7						
QY	1822	tgattc 1827							
Db	6	TGATTC 1							

FEATURES		High quality sequence stop: 849.		Location/Qualifiers	
source		1..864		/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4864554" /clone_1ib="NIH_MGC_15" /tissue_type="adenocarcinoma cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."	
BASE COUNT		235 a		204 c 210 g 215 t	
ORIGIN					
Query Match		29.8%; Score 561.2; DB 10; Length 864;			
Best Local Similarity		96.7%; Pred. No. 4.1e-152;			
Matches		594; Conservative		0; Mismatches 18; Indels 2; Gaps 2;	
QY	1	gattgtctccctccacaccttccctgtgtcctgagctccacacctgtcccaattgtgaag	60		
Db	252	GATTGTCTCTCTCCACACCTTCCCTGTGCTGCGGTCCACACTGTCTCCACTCTGTGAGCG	311		
QY	61	atgttcaatgga-agaagctctgcagctgcaacttctgttggtctgtgctgtat	119		
Db	312	ATGTTCAATGGAAGAGAGATCTGCAGCTGCATTAATCTGTGGCTGTGGCTGTATAT	371		
QY	120	gtctgtgcacactgtgtctctgaacttctcttcaaggttgaagtgtgacttgacacct	179		
Db	372	GCTGTGCGCAGCTGTGCTCTGAACCTTCTTCAAGTTGAAGTGTGACTGTACACCAT	431		
QY	180	gggtctggaggtccaggggaatcctaagcagatcgttaggaatatttgtataattct	239		
Db	432	GGGTCTGGAGTCCAGGGAATCTCAAGCCAGTCTGTAGGAATCTGTATTAATTTCT	491		
QY	240	gaacttccagaaagaaggtcttcaactgttcaagggttccacggagggagcaagaagc	299		
Db	492	GAAACTTCCAGCAAGAGGCTTATCACTGTTTAGGGGTACCCGAGGGACCAAGAGC	551		
QY	300	agtgctcaggtatctgaataaccttgagaggtcagaagaagcgagagccttccacaga	359		
Db	552	AGTGCTTCAAGGCTATTCTGATTAACCTGGAGGTCAAGAAAGCAGAGGCTTTCACAGA	611		
QY	360	caccacactaccttccctccacagagactgtgagcaacttcaaggtgaaagaagttcat	419		
Db	612	CACCCACTACCTTCCCTCACAGAGACTGTGAGCACTTCAAGGCTGAAGGAAGTTCAAT	671		
QY	420	acagttcccaacttgagaagaagaaggttgagttccctattgtatctatcgttgatt	479		
Db	672	ACAGTTCCCACTGAGCAAGAGAGGTGAGTCCCTATTGCACTTATGTGATTTA	731		
QY	480	tgaagaagtgaagaacttgaagaagctactgagctgtgtatgcccctcaagaacatata	539		
Db	732	TGAGAAGATTGAAACTTTGAAAGGCTACTGCAGCTGTGTATGTGCCCTCAGAACATTA	791		
QY	540	ctgtgtccatgttgatgaagaagtcaccagaacttcaagaagcggttcaagaactat	599		
Db	792	CTGTGTCCCTGTGTGATGAGAGT-CCAGAACTTCCAAAGCAGCGCAAGCAATTAAT	850		
QY	600	ttctgtctccca	613		
Db	851	TTCTGTGTCCAAA	864		

Search completed: September 26, 2002, 02:55:07

Job time: 3832 sec

Search completed: September 26, 2002, 02:55:07
Job time: 3852 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:23:02 ; Search time 4629.8 Seconds
(without alignments)
8515.615 Million cell updates/sec

Title: US-09-874-390-1_COPY_436_2319
Perfect score: 1884
Sequence: 1 gattgtgtctctctccacct.....ataataaatacctaacaga 1884

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_da:*
- 2: gb_hcg:*
- 3: gb_in:*
- 4: gb_ow:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_pl:*
- 8: gb_pr:*
- 9: gb_ro:*
- 10: gb_ry:*
- 11: gb_sy:*
- 12: gb_un:*
- 13: gb_vl:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_uts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hcg_hum:*
- 31: em_hcg_inv:*
- 32: em_hcg_other:*
- 33: em_hcgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	ID	Description

1	1880.8	99.8	187275	2	AC092755	AC092755 Homo sapi
2	1833	97.3	2106	9	BC017032	BC017032 Homo sapi
3	1829.8	97.1	2216	6	AX045249	AX045249 Sequence
4	1818.8	96.5	2217	6	AX045251	AX045251 Sequence
5	1818.8	96.5	2217	6	AX045253	AX045253 Sequence
6	1798.8	95.5	2128	6	AF136078	AF136078 Sequence
7	1798.8	95.5	2128	6	AF102542	AF102542 Homo sapi
8	1317	69.9	1317	6	AF038650	AF038650 Homo sapi
9	1317	69.9	1317	6	AF038650	AF038650 Homo sapi
10	1011	53.7	108873	14	AF318573	AF318573 Bovine he
11	1001.4	53.2	2017	14	AF231105	AF231105 Bovine he
12	960	51.0	4179	14	BHVARBEG	BHVARBEG Bovine he
13	923.8	49.0	1841	10	BC018297	BC018297 Mus muscu
14	726.2	38.5	176483	2	AC096432	AC096432 Rattus no
15	376.8	20.0	2033	10	MMU19265	MMU19265 Mus muscu
16	375.2	19.9	2023	10	D87333	D87333 Mus muscu
17	375.2	19.9	2047	10	D87332	D87332 Mus muscu
18	374.2	19.9	377	6	AX341786	AX341786 Sequence
19	369.4	19.6	1287	6	AX087946	AX087946 Sequence
20	369.4	19.6	2105	6	I41262	I41262 Sequence 3
21	369.4	19.6	2105	6	I61440	I61440 Sequence 3
22	369.4	19.6	2105	6	I72548	I72548 Sequence 3
23	369.4	19.6	2110	6	AX333255	AX333255 Sequence
24	369.4	19.6	2110	6	AX335500	AX335500 Sequence
25	369.4	19.6	2110	6	HUMGCGG	HUMGCGG Human beta-
26	369.4	19.6	2204	9	HUMGCGG	HUMGCGG Human beta-
27	367.8	19.5	159930	2	AL138757	AL138757 Homo sapi
28	367.8	19.5	183858	9	AL161626	AL161626 Human DNA
29	364.6	19.4	5010	10	S79797	S79797 enzymatic 9
30	364.2	19.3	2435	4	BTU41320	BTU41320 Bos taurus
31	307.2	16.3	309	6	AX341869	AX341869 Sequence
32	305.6	16.2	309	6	AX340693	AX340693 Sequence
33	257	13.6	380	14	HSBROVHEZF	HSBROVHEZF Bovine he
34	243.2	12.9	3083	9	HUMB16NACT	HUMB16NACT Homo sapien
35	243.2	12.9	151136	9	HSAB19K21	HSAB19K21 Human DNA
36	207.8	11.0	189534	2	AC095187	AC095187 Rattus no
37	203	10.8	252504	2	AC073815	AC073815 Mus muscu
38	198.8	10.6	155359	9	AL358777	AL358777 Human DNA
39	198.8	10.6	166788	2	AC021626	AC021626 Homo sapi
40	191.8	10.2	1362	6	AX087935	AX087935 Sequence
41	191.8	10.2	3435	9	AF132035	AF132035 Homo sapi
42	189.6	10.1	3508	10	AB037596	AB037596 Mus muscu
43	188.6	10.0	184590	2	AC093259	AC093259 Homo sapi
44	185	9.8	63853	2	AC087663	AC087663 Homo sapi
45	184.4	9.8	189534	2	AC095187	AC095187 Rattus no

ALIGNMENTS

RESULT 1

AC092755 187275 bp DNA linear HTG 19-JAN-2002

LOCUS Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***

DEFINITION AC092755 AC022480

ACCESSION AC092755.3 GI:18249989

VERSION HTG: HTGS_P18249989

KEYWORDS HTG: HTGS_P18249989; HTGS_ACTIVEIN.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Burke, J., Dots, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D. and Hood, L.

TITLE Sequencing of human chromosome 15 D15S146-D15S117 region

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 187275)

AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Burke, J., Dots, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R., Pate, D. and Hood, L.

TITLE Direct Submission

Dh	789	TGtGTCCATGTGGATAGAGAGTGTCCCGAAGACTTTTAAAGAGCGGTCAAGCAATTTATT	848
Oy	601	tcttgctccccaatgltcttcataagccagtaagctggtctcgggtggtttatgctcctctg	660
Dh	849	TCTTGCTTCCCAATGTCTTCATAGCCAGTAACCTGGTTCGGGTGGATTATCTCTCTGG	908
Oy	661	tccagggtgcaagtgagccctcaactgcatggaagactgtgtccagagctcagtgccgtg	720
Dh	909	TTCAGGGGTGCAAGGTGACCTCACTGATGGAAGACTTCTCCAGAGCTCACTGTCCGGG	968
Oy	721	aaatactctccgaaatacatgtggaagagacttccctataagcaatgcaagaatgctc	780
Dh	969	AAATRACTTCCGATPACATGTGGAGCGACTTTCCTATAAGAGCAATGACAGATGGTC	1028
Oy	781	caggctcctaagatgtgtgaatgggaatagacatgagtcagaggttaactcctaagcac	840
Dh	1029	CAGGCTCACAAGATGTGAATGGAGAGAAATAGCATGGAGTCAAGAGTACTCTTAAGAC	1088
Oy	841	aaagaacccgctgynaatatcaacttgagtgatgagagacacatatacctaaccac	900
Dh	1089	AAAGAACCCCTGGAAATPCTACTTTGAGAGTGTAGAGACACATTTACACTTAACAC	1148
Oy	901	aagaagaaggtctctccccccttaatttaacatgttttaagggaatgcygacattgtg	960
Dh	1149	AAGAGAAGAGATCTCTCCCTTTAAATTTAACTATGTTPACAGGGAAATCCGACATTTGTG	1208
Oy	961	gcttcccgagatttcgcgcacaatggttttgaagaaccttaatcccaactgtgtga	1020
Dh	1209	GCTTCCGAGATTTGTCGCAACATGTTTGAANAACCTTAATCCCAACACTGATTGA	1268
Oy	1021	tgggtlaaagaacacttataagccagatgaacacctctgggcaaccttaagcgtgcag	1080
Dh	1269	TGGGTAAAAGACACTTPTATACCCAGATGAACACTCTGGGCCACCTTCAGCGTGCACGG	1328
Oy	1081	tggatgcctgtgctctgtctccacaacccccaaatgaagaattcgaagatgctctatt	1140
Dh	1329	TGGATGCTGTGCTGTGTCCAAACACCCCAATPACAGATCTCAAGACTGTGACTTTAT	1388
Oy	1141	gcacagctgtygcaagtgycagaggttcaatgaagagacatgataaggtgtcctctatg	1200
Dh	1389	GCCAGCGGTGCAAGTGGCAGAGGTCATGAGGGAGACATCGATAAGGTGTCTCTTATGCT	1448
Oy	1201	ccctgctctggaatccacagcagggtcatalctgcgtttatggggctggggacttgaattg	1260
Dh	1449	CCCTGCTCTGTGAATCCACACGAGCGGGCTATCTGCGTTATAGGGCTGGGGACTTGAATGG	1508
Oy	1261	atgcttcaaaacacataccgtgtggccacaagtttgaccocaaagtatgatatgct	1320
Dh	1509	ATGCTCTAAAACCACTACCTGTTGGCCACAAATTTGACCCAAAGGTAAATGATATATCT	1568
Oy	1321	cttcagtgcttagaagaatacctacgtttataaagccatctatggaagcttaacttgaagc	1380
Dh	1569	CTTCAGTGCTTAGAGAAATACCTACGTATTAAGGCCATGTATGGGACTTAACATTGAAAC	1628
Oy	1381	acacataigaagcgtgtgctacactgtggyggaagagacatgtacaaacatgctcagaactg	1440
Dh	1629	ACACTATGAGAGCGTGTCTACTCTGTGGGGCAAGACATGTACAAACATGCTCAGAACTTG	1688
Oy	1441	cttggaagcagtgtyggtggygagaccagagcttggaattcgtgycatccttaagataa	1500
Dh	1689	CTTGGAACAGTGTGGTGTGGAGACCAAGGGCTTTCATTTGTGGCAATCTTTTAGGTAAGA	1748
Oy	1501	ggagctgcattagaatgtggtgtaagaagatcttgctccttgaaatgtctgctbgygga	1560
Dh	1749	GGGCTGCTATTAGATTGTGGGTAAAGATCTTTTCCCTTGCAAAATGTGTGCTGGGTGA	1808
Oy	1561	atgctgctgttctctlaaccccctaaccctagtagtctccctcaactaacttctcctaagt	1620
Dh	1809	ATGCTGCTGTGTTCTCAACCCCTTAACCCCTAAGTAGTTCTCTCCACTAATCTTCTCACTAGT	1868
Oy	1621	gagaaatagaactgctgtgtataaggaagatgaaagaggatagtggtataagcactgtat	1680
Dh	1869	GAGAAATAGAACTGCTGTGTATAGGAGAGTGAAGAGGGATTTGTGTAGAGCACTTGAT	1928

QY	1681	ttcagttgaatgctcgtcgtgtgagcttccattcgttgagctgcgcgttccataatc	1740
QY	1681	ttcagttgaatgctcgtcgtgtgagcttccattcgttgagctgcgcgttccataatc	1740
DB	1929	tttcagttgaaatccgcgtgtagcttcttccattcgttgagctgcgcgttccataatc	1988
QY	1741	caggttcgtgaacgtgaggaagaaacttggatggaagaagaacctcccttcgtactgt	1800
DB	1989	caggttcgtgaacgtgaggaagaaacttggatggaagaagaacctcccttcgtactgt	2048
QY	1801	aacttaaaataaatagctcctgattcaagta	1833
DB	2049	aaacttaaaataaatagctcctgattcaagta	2081
RESULT	3		
AX045249	AX045249	2216 bp	DNA
LOCUS	Sequence	1 from Patent WO0066727.	linear
DEFINITION	AX045249		PAT 24-NOV-2000
ACCESSION	AX045249.1	GI:11343799	
VERSION			
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2216)		
TITLE	Adolf, G., Heider, K. H. and Sommergruber, W.		
JOURNAL	Tumour-associated antigen Patent: WO 006727-A 1 09-NOV-2000; Boehringer Ingelheim International GmbH (DE)		
FEATURES	Location/Qualifiers		
source	1..2216		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
5'UTR	1..426		
CDS	427..1743		
	/note="unamed protein product"		
	/codon_start=1		
	/protein_id="CAJ1734.1"		
	/db_xref="GI:11343800"		
	/translation="MVGWKRRLQQLHYLMAAGCYLLATVALKLSFRLKDSHGLEESS		
	RESOYCRNLYNFELKLPARKSINGSGVGVGDEAVLQALINLVKKKKEPTDTH		
	YLSLRDCEHFKAEKREIQEPLEKEEVEPLAYSMVHEKLENERELRLAFAPONTY		
	CVHDEKSPERFKAEKVAIKFISCPNVAFLAVVVAASMSRVOADINCMEDLQSSV		
	PKKFLNMGCDPRKISNAEMVQALMKINGNSMESVPPKHETRRKYHREYVRLTL		
	HLTKKKDDPPRYNLTMFTGNAYIASKDPVOHYLKNKRSQQLIEWKDYTSFEDLMA		
	TLQKRAKPGSVPNHPRYDISDMYSIAKLVMQGHEDIDGAPYAFCSGIHORALCV		
	YGADDLMMMLNHLNLFKEDPKVDNALQCLEEYLRKALYGTFL		
3'UTR	1744..2216		
BASE COUNT	600 a 499 c 547 g 570 t		
ORIGIN			
Query Match	97.1%; Score 1829.8; DB 6; Length 2216;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1831; Conservative	0; Mismatches 2; Indels 0; Gaps 0;		
QY	1	gattgtgtcctctccaccctcctcctgtcgtcgtggttccacgtgttcccatctgtgacg	60
DB	367	GATTGTGACTCTCTCCACACTTCCCTGCTCGTGCCTTCACACGTCTCCCATTTCTGTGACG	426
QY	61	atggttcaatgaagaagactctgccaagctgatactactgtgtgagcttgagctgtatag	120
DB	427	ATGGTTCATGTGAAGAGACTCTGCCAGCTGCATTACTGTGTGGCTCTGGCTGCTATATG	486
QY	121	ctgctggccactgtgtgacctgaaactcttcttcagtgatgaaagtgtactctgacacattg	180
DB	487	CTGCTGGCCACTGTGTGACCTGAAACTTCTTTCAGGTGGAAGTGTGACTCTGACCATTTG	546
QY	181	ggtctggagctcagggaatctcaaaagcaagctacgttaggaatactctgtataattcctg	240
DB	547	GGTCTGGAGCTCCAGGGAATCTCAAAAGCCAGTACTAGGAATATCTTGTATTAATTTCTCG	606

Oy 241 aaattccagcaagaaggtctatcaactgtgtcagggtgtaaccgagggaaccaagagca 300
 |||||
 Db 607 AAATCTTCAGCAAGAGAGGTCTATCAACTGTTCAGGGGTCACCCGAGGGGACCAAGAGGA 666
 Oy 301 gtgtctcagggtactcttgaataactgtgaagtcgaagaagcgagagcttccacaagc 360
 |||||
 Db 667 GGGCTTCAGGGCTATCTGATTAACCTGGAGGTCAAGAGAACGAGAGCTTCACAGAC 726
 Oy 361 acccaactaccctccctcaaccagagactgtgagcacttcaagctggaagaagtaactata 420
 |||||
 Db 727 ACCCACTACTCTCCCTCCACCGAGACTGTGACACTTCGAAGGCTGAAAGAGATTCTATA 786
 Oy 421 caattcccaactgagcaagaaggtggagttccctattgtcaactctatgtatgtatcaat 480
 |||||
 Db 787 CAGTTCCCACTGAGCAAGAGAGGTGGAGTCCCTATTGCATACTGATGTGATTAAT 846
 Oy 481 gagaagatgtgaaacttggaaagctactgtcgagctgtgtatgtccctcagaacataac 540
 |||||
 Db 847 GAGAAGATTGAAACCTTGGAAAGGCTACTGCGAGCTGTGTATGCCCTCAGAACATATAC 906
 Oy 541 tgtgtccatgtgtgagagaagtcgcccaagaactttcaagaagcggtcgaagaacttatt 600
 |||||
 Db 907 TGTGTCCATGTGTGATGAGAGAGTCCCAAGAACTTTCMAAGAGCGGCTCAAGCAATTATT 966
 Oy 601 tctgtctcccaaatgtctctcaagctagccagtaagctgtgttcggtgtgtatgtctctc 660
 |||||
 Db 967 TCTTGTCTTCCCAATGTGTCTTATAGCCAGTAAGCTGTGTTCCGGGTGTTATCCCTCTGG 1026
 Oy 661 tccaggtgtgcaagctgtgacctcaactgtcagtaagactgtgtccagagctcagtgccg 720
 |||||
 Db 1027 TCCAGAGGTGCAAGCTGACCTCAACTGACGTAAGACTTTCCTCAAGAGCTCACTGCCGTGG 1086
 Oy 721 aaatacttcctgaatacatgtgtgagcgagacttccctcaagaagaactcagaagtggtc 780
 |||||
 Db 1087 AAATACTTCTCAATATCATGTGTGGACGACTTTCCTATTAAGACCAATGCATAGATGCTC 1146
 Oy 781 caggtctcctaagaatgttgaatggagaatagacatgagatcagagtaactcctaagcac 840
 |||||
 Db 1147 CAGGCTTCTCAAGATGTGTGAATGGAGAAATATGATGAGATGATGATGATGATGATGATG 1206
 Oy 841 aagaagaacccgtgtgaaataatacttgaagtgatgtgagagaacattatacctaacaac 900
 |||||
 Db 1207 AAAGAAACCCGCTGGAATAATCATCTTGAAGTAGAGACACTTAAACCTTAACCAAC 1266
 Oy 901 aagaagaagatcccccctataatctaactatgtttacaggggaatgcgtacattgtg 960
 |||||
 Db 1267 AAGAAGAAGATCTCCCTTATATTTAACTATGTTCACGAGGATTCGATCATTTGG 1326
 Oy 961 gcttcccgagatltcgtccaacatgttltgaagaacccctaatacccaactgtattga 1020
 |||||
 Db 1327 GCTTCCCGAGATTGTGTCACACATGTTTTGAAGAACCCCTAATCCCAACACTGATTTGA 1386
 Oy 1021 tgggtaaagaacactatagcccaagatgaacacctcttgggccaaccttcaagtgtaag 1080
 |||||
 Db 1387 TGGGTAAAGACATTATAGCCCAATGAAACCTCTGGGCCACCTTACAGCTGCACAG 1446
 Oy 1081 tggatgtcctgtgtctgttcccaacaccccaagtaagacatctcaagatacttctatt 1140
 |||||
 Db 1447 TGGATGCCGCTGCTGTTCACCAACCCCAAGTACGACATCTCAGACATGACTTCTATT 1506
 Oy 1141 gccagagctgtgtcaagtggtcaggtgtcatgaggagacatgataaggtgtctcttattgt 1200
 |||||
 Db 1507 GCCAGGCTGTGCAAGTGGGAGGTCATGAGGAGACATGATGAAGGTGCTCTTATGCT 1566
 Oy 1201 cccgtctgtgaatacccaacgaggtctatgtgttattgaggtctggggaactgtgaattgg 1260
 |||||
 Db 1567 CCTCTCTGGAATCACACAGGGGCTATCTCGTTTATGGGGCTGGGACTTGAATTGG 1626
 Oy 1261 atgtcttcaaacatcatcctgtgtgccaagaagttgacccaagaagtatgatataatgt 1320
 |||||
 Db 1627 ATGCTTCAAAACCATCACTGTTGGCCAAAGTTTGAACCCAAAGGTAGATATATGCT 1686
 Oy 1321 ctctagtgcttagaagaataactacgtataagccatctatgagactgaacttgagac 1380

Db 1687 CTTACGTGCTTAGAAGAAATACCTAGTTATAGGCCATCTATGAGACTGAACCTTGAGAC 1746
 |||||
 Oy 1381 acacatagagacgttgtctactctgtgtggcaagacatgataaacatgtctagaacttg 1440
 |||||
 Db 1747 ACACATATGAGACGTTGCTACTGCTGGGCAAGACATGATACAAACATGCTCAGAACTTG 1806
 Oy 1441 ctggagacagttgtgtgtgagaccaggtcttgcaatcctgtgcatcccttagagaaga 1500
 |||||
 Db 1807 CTGGGACACTGTGGGTGGAGACAGGCTTTGCATTTCGTGTGATCCTTTAGATTAAGA 1866
 Oy 1501 gggctgtattagattgtgtgtaagtagatccttgccttggcaaatgtcgtcgtgtgtga 1560
 |||||
 Db 1867 GGGCTGCTATTATGATTGTGGTAAGTAGATCTTGTGCTTCAAAATGCTCGCTGGGTGA 1926
 Oy 1561 atgtcgtctgttctctcaacccttaacccttagtagtctcctcaactcttccactagt 1620
 |||||
 Db 1927 ATGCTGCTTGTCTCCACCCCTTAACCTAGTAGTTCCTCCACTACTTTCACATTAAGT 1986
 Oy 1621 gagaatgaaactgtctgtatagagagatgtgaagagagatatgtgttagagacttgat 1680
 |||||
 Db 1987 GAGATGAGAACTGCTGTGTATAGGAGACTGAAGAGAGATATGTGTAGACACTTGAT 2046
 Oy 1681 ttcagltgaatgctgt 1740
 |||||
 Db 2047 TTCAGTTGAAATGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2106
 Oy 1741 caggt 1800
 |||||
 Db 2107 CAGGTTTGT 2166
 Oy 1801 aacttaaaataatagctcctgattcaagta 1833
 |||||
 Db 2167 AACTTAATAATAATAGCTCCTGATTCAAGTA 2199
 |||||
 RESULT 4
 AX045251 2217 bp DNA linear PAT 24-NOV-2000
 LOCUS AX045251
 DEFINITION Sequence 3 from Patent WO0066727.
 ACCESSION AX045251
 VERSION AX045251.1 GI:11343801
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 2217)
 AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
 TITLES Tumour-associated antigen
 JOURNAL Patent: WO 0066727-A 3 09-NOV-2000;
 Boehringer Ingelheim International GmbH (DE)
 FEATURES
 source
 1..2217
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 5'UTR 1..844
 3'UTR 1..426
 CDS 427..999
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17335.1"
 /db_xref="gi:11343802"
 /translation="MVKRRICQLHYLMALGCMYLATVAIKLSFRLKCDSDHGLGES
 RESOSQICRNILVFLKLPKAWSTNGSVTRGDEAVLQALINLEVKKKREPTDTH
 YLSLRDCEHFKARKEIOPFLSKSEVEFFPLAYSMVIEHKIENRRLRAYAPONTY
 CVHDEEVPNRFGSGSNYFLPKLHSQ"
 845..1744
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17336.1"
 /db_xref="gi:11343803"
 /translation="MRRILTKLGYCELCLMPRTYTVSMNKKSPDFEAVKAIISCF

PNVFIAKLVRYVYASRKYOADLNCMEDLLQSSVEMKFLANTCGDPEPKSAEMQ
ALMLNGRNSMESEVPKHKETRMKYHFEVVRDPLHLTNKMKPPNLTMPGNAYI
VASRDEVOHYLVKNPKSQOULEWYKDYSPLEHMLTORARMKPGSVNHPKXDIDM
TSIARLVKMGHEGIDIDKGAPAPCSGIHQRAICVYGAGDNLMLNHHLLANKFDPK
VDNALOCLEERYARKAYGTEL"
3'UTR 1000..2217
3'UTR 1745..2217
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match 96.5%; Score 1818.8; DB 6; Length 2217;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 gatctgtccctcccaactccctgtctgtctgtccacactgtccacactgtgacg 60
Db 367 GATTGTGACTCCCTCCACACTTCCCTGTGCTGCTGCTCACCCTGTCCCATTTCTGTGACG 426
Qy 61 atgttcaatggaagagactctgcacgtcactactgtgtgtgtctctgtgtctata 120
Db 427 ATGTTCAATGGAAGAGACTCTGCCACTGATTAATTGAGGCTCTGGCTCTATATG 486
Qy 121 ctgtgtgcaactgt 180
Db 487 CTGCTGACCACTGTGCTGTGAACCTTTCAGTTGAAGTGTGACTGTGACCTGACACTG 546
Qy 181 gctctgagtcacaggaatctcaaaagcagctactgttagaatactgtataatctctg 240
Db 547 GGTGTGAGTCCAGGAATCTCAAAAGCAGTGTAGAAATATCTGTATATTTCTG 606
Qy 241 aaacttcagaagaagagctctataactgttcaagggttcacgcgaaggagaccaga 300
Db 607 AAATCTTCAGAGAAAGTGTGTATACACTGTTCAGGGGTCAACCGAGGGAGCAAGAGCA 666
Qy 301 gtgtcttagcactctctgaataactgtgaggttcaagaagaagagagcctttacagac 360
Db 667 GTGCTTACAGCTATTTGATTAACCTGAGGTCAAGAGAGAGCAGAGACCTTTTACAGAC 726
Qy 361 accactactctctcccaacagagactgttagacacttcaaggtctgaagaagaatcata 420
Db 727 ACCCACTACCTCTCCCTACACAGAGACTGTGAGCACTTCAAGGCTGAAGAGACTTCATA 786
Qy 421 cagttcccaactgagcaagaagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
Db 787 CAGTTTCCACACTGAGCAAGAGAGGTGAGTTCCTATTGATCACTCATGCTGATTCAT 846
Qy 481 gagaagaatgaaaacttgaagaagctactgtgagctgtgtgtgtgtgtgtgtgtgtgtgt 540
Db 847 GAGAGATTTGAAAACCTTGAAGAGCTGAGACTGTGTATGCTCCCTCAGAACATATAC 906
Qy 541 tgtgtcactgtgagtg-agaagtcaccagaacttcaagaagggcggltcaagaactat 599
Db 907 TGTGTCCATGATGAGAAAGTCCCAAGAACTTTCAAAGAGCGGTCAAGCAATATAT 966
Qy 600 ttctgtgtcccaatgtcttccatagccagttaagctgtgtgtgtgtgtgtgtgtgtgt 659
Db 967 TTTCTGTGTCCCAATATGTTTCATATGACAGTGTGAGTGTGAGTGTATATGCTCTGTG 1026
Qy 660 gtccagaggtgcaagctgacactcaactcatggaagactgtgtccagagctcagtgccgt 719
Db 1027 GTCCAGGGTGAACCTGACCTCACTGATGGAAGACTGTGTCAGAGCTGAGTGTGCTG 1086
Qy 720 gaaatctctcgaatacatgtgtggaagcgaacttccataaagaagcaatgcaagatgt 779
Db 1087 GAATTAATCTCTGTAATACATGTGGAGAGACTTTCATATAAAGCAATGACAGATGAT 1146
Qy 780 ccaaggtctcaagaatgttgaatggaggaataagcatggaatcagaagtactcctaaga 839
Db 1147 CCAGAGCTCTCAAGATGTTGATGGAGGAATGAGCATGAGTCAAGAGTACTCTTAAGCA 1206
Qy 840 caagaagaaccgctggaataatactttagaggttagagagaacattacccaacaacaa 899

Db 1207 CAAAGAAACCCGCTGGAATAATACCTTTGAGTGTAGAGACACATTACACTAACCA 1266
Qy 900 caagaagaagatccctcccttaataatcaactatgtttcaagaagatgcatatgt 959
Db 1267 CAAGAAAGAGATCTCCCTTATATTTAATATTTAATTTACAGAGGAATGATGATTTGT 1326
Qy 960 ggtctccgagatctgttccacaatgtttgaagaacccataatccacaactgtatga 1019
Db 1327 GGTCTCCGAGATTTCTCCACACATGTTTGAAGAACCTTAATCCCAACACAGATGA 1386
Qy 1020 atgggttaagaagacactatagcccaagatgaacactcttggtccaccccttaagctgac 1079
Db 1387 ATGGGTAAAGACACTTATAGCCAGATGACACCTGTGGCCACCTTACGCTGACG 1446
Qy 1080 gtgatgcctggtctgttcccaacaccccaagtaagacatctcagacatgtactcat 1139
Db 1447 GTGATGCTGCGTCTGTTCACAAACCCCAAGTACGATCTCAGACATGATTTCTAT 1506
Qy 1140 tgcagagctggtcaagtgtgcaaggttcaatgaggaagacatgataaagggtctctatgc 1199
Db 1507 TGCAGAGCTGCTCAAGTGGCAGGGTCAATGAGGGAGACATGATTAAGGCTCTCTTATGC 1566
Qy 1200 tccctgtcttgaatgcacagcggtctatctgtttatggtgtgtgtgtgtgtgtgtgtgt 1259
Db 1567 TCCCTGCTTGGAATCCACAGCGGGCTATCTGCTTTATGGGCTGGGACTTGAAATG 1626
Qy 1260 gatgcttcaaaaccatcacctgt 1319
Db 1627 GATGCTTCAAAACCATCACCTGTGGCCCAACAAATTTGACCAAGGATGATGATATGC 1686
Qy 1320 tcttcaagttgttgaagaataactaagtttaagggcatctatagggactgaactttgaga 1379
Db 1687 TCTTCAGTGTCTTGAAGAAATACATCACTGTATTAAGCCATCTATGGACTGAACCTTAGA 1746
Qy 1380 cacactatgagagcgt 1439
Db 1747 CACACTTGAAGAGGCTGTGCTACCTGTGGCCAGAGCATGTACAAATGCTCAGAACTT 1806
Qy 1440 gcttggaacagtggt 1499
Db 1807 GCTGGGACAGTGTGGTGGGAGACAGGCTTTGCAATTTGTGTGGCATCTTATAGGATAG 1866
Qy 1500 aggtgtctatgttgaatgt 1559
Db 1867 AGGGCTCTATTAATGTTGGGTAAAGATCTTTCCTTGCATAATTTGCTGTGGGTG 1926
Qy 1560 aatgtgtctgttctccacaccccaacccatagatgttccctccacacttctcactaag 1619
Db 1927 AATGCTGCTGTCTCTCACCCCTTAACCTTAAGTGTCTCCACTTCTCAGTAAG 1986
Qy 1620 tgaagaatgaagactgtctgtataggaaggtgaagaaggataltgtgttagagcaattga 1679
Db 1987 TGAATAATGAGAACTGTGTGTATAGGAGAGTGAAGGAGGATATGTGTAGAGCACTTGA 2046
Qy 1680 tttaagttgaagctgt 1739
Db 2047 TTTAGTGTGAATGCTCTGTGTATCTTTCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2106
Qy 1740 ccaaggttgttgaaggttgaagaagacttgaatggaagaagaacacttccctctgtactgt 1799
Db 2107 CCAGGTTTGGTAGCGTGAGAGAGAACTTGTATGAAAGAGAACTTCCCTCTGTACTGT 2166
Qy 1800 taacttaaaataaataagctctgtatcctaagta 1833
Db 2167 TAACTTAAAAATAATAGCTCTGTGATTCAAAAGTA 2200

RESULT 5
AX045253 2217 bp DNA linear PAT 24-NOV-2000
LOCUS AX045253
DEFINITION Sequence 5 from Patent WO006727.
ACCESSION AX045253
VERSION AX045253.1 GI:11343804

OY	1680	ttcaagtgtgaatgcctgctcgtgtaagcttccattcctgtggagctgcgcgttcctaataatt	1739
Dd	2047	TTTTCAAGTTGAAGAAGCCCTGCATGCGTAGCATTTCACATTTGTGGTACCCTGCCGTTCCATAAAT	2106
OY	1740	ccaagtttgtagtcgtlgagaaggaaacttgtatcgaaaagaacattccccttcctgtacct	1799
Dd	2107	CCAGGTTTGTAAGCTGTGGAGAGAACTTTGANTGAAAAGAACCTTCCTCTGTACTGT	2166
OY	1800	tactttaaaaaaatagctccctgatctccaagta	1833
Dd	2167	TAACTTAAAAATAAATAGCTCCTCATTTCAAAGTA	2200
RESULT	6		
ARI36078			
LOCUS	ARI36078	2128 bp	DNA linear PAT 16-JUN-2001
DEFINITION	Sequence 1 from patent US 6136580.		
ACCESSION	ARI36078		
VERSION	ARI36078.1 GI:14476750		
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2128)		
TITLE	Fukuda,M. and Yeh,J. .beta.-1,6-N-acetylglucosaminyltransferase that forms core 2, core 4 and 1 branches		
JOURNAL	Patent: US 6136580-A 1 24-OCT-2000;		
FEATURES	Location/Qualifiers		
source	1..2128		
BASE COUNT	569 a 477 c 526 g 556 t		
ORIGIN	/organism="unknown"		
Query Match	95.5%; Score 1798.8; DB 6; Length 2128; Best Local Similarity 99.8%; Pred. No. 0; Matches 1812; Conservative 0; Mismatches 2; Indels 2; Gaps 1:		
OY	1	gattgtgccctccaccacctccctcgtgctggctggctccacctcgtctcccatctgtgacg	60
Dd	294	GATTGTGTCCTCCTCACACCTTCCTCTGCTGCTGCCTGCCACTGTCTCCCATTTCTGTGACG	353
OY	61	atggttcaaataagaagactctccacgtcacatctactgttgscgtctggcgcgtatatg	120
Dd	354	ATGGTTCAATGGAAGAAGACTCTGCCAGCTGCATTACTGTGGGCTCTGGGCTGCTATATG	413
OY	121	ctgcgcgcacactgtgctctgtaaactttcttcacagttgaagtgtgactctgacaactgy	180
Dd	414	C TGCTGGCCAC TGTGGCTCTGA AACTTTCTT TCAGGTTGA ATGTG ACTCTG ACCACTTG	473
OY	181	ggctcgtgagatccagggaaatctcaaacccagtatctgtagaatatcttgtataattctcy	240
Dd	474	GGCTCGGAGATCCAGGGAAATCTCAAGGCCAGTACTGTAGGAATAATCTTTGATTAATTCCCG	533
OY	241	aacattccagaaaagaagfctaaactaaactgttaagggttcaaccggagggaccaagaagga	300
Dd	534	AAACTTCCACGA AAAGG GTCTATCAACTGTTT AGGGG GTCA CCGGAG GGAC CAAGA GGCA	593
OY	301	gtcgttcaggtcatctcgaataactgtaggtgcaagaagaacgagagacctticacaagac	360
Dd	594	GTCGTTCAAGCGTATTTCTGAATAACCTGGAGGTCMAAGAAGAACAACGAGGCTTTCACAGAC	653
OY	361	accacatcctctccctcaacagagactgtgagcaactcaaggctgaaagaagtlcata	420
Dd	654	ACCCTACTACTCTCCCTCCACAGACACTGTGAGCACTTCAAGGCTGAAGAAGAGATTCAATA	713
OY	421	cagttcccaactgagaagaagaaggtlgaagttccctatttgataactctatgtgtattcat	480
Dd	714	CAGTTCCCACTGAGAAAGAAGGTTGGAGTCCCATTATTCCTTAATTCATTATGTGTATTCAAT	773
OY	481	gagaagattgaaaacttbaaagagctactgcagactgtatagcccccttaagaacatatc	540

Db	774	GAGAAAGATTGAAACCTTTGAAAGGCTACTCGAGCTGTGTATGCCCCCTCAGAACATATAC	833
QY	541	tggtccatgtgtgatgtgaagtcgccagaaacttcaaagagcggltcaagaactatt	600
Db	834	TGTGTCCATGTGTGATGTGAAGTCCGCCAGAAACTTTCAAAGAGCGCGTCAAAACATTTAT	893
QY	601	tctgtcttcccaaatgtcttccatagacgaagaaactgtgttcgggtgtgttatgtccctccg	660
Db	894	TCTTGTCTTCCCAAAATGTCTTCATATGCCAGTAACTGTGTGGGTGGTATATGCTCTCCGG	953
QY	661	tccaagtggtcaagcttgacccctaactgcaltggaagacttgtctccaagctcagtgccgtg	720
Db	954	TCCAGGGGTGCAAGCTGCATCTCACTGATGGAAGACTTCTCTCCAGAGCTCAATGCGCTGG	1013
QY	721	aaatcttccatgaatacatgttgggaagggacttccctataaagacaatgacagatgtc	780
Db	1014	AAATCTTTCCTCGAATACATGTGGGACGGACTTTCCTATTAAGCAATGTCAAGATGTGTC	1073
QY	781	caggctctcaagatgttctgaatgggaagatagatgtgaatcagtgacgttaactctaaagac	840
Db	1074	CAGGCTCTCAAGATGTGTGATGGGAGGAATAGCATGAGATCTCAGAGTACCTCTAAGAC	1133
QY	841	aaagaacccgcgtggaaatatcaactttaggtagtgagagacacataaccttaacoac	900
Db	1134	AAAGAAACCCCTCGAATATCACTTTGAGGTGTAGAGACATATACCTTAACCTACAC	1193
QY	901	aagaagaagagctctccccccttaatttaactcatttcaagggaaatgcgtacatttg	960
Db	1194	AAGAAAGAAAGATCCTCCCTTATATTTATTTACATATGTTTACAGGGAATGCGTACATTGTG	1253
QY	961	gtctccagaatcttcgtccaacabgttttgaagaacctaaatcccaaacatgtatga	1020
Db	1254	GCTTCCCGAGATTTCTGTCACAACTGTTTTGAAACCTTAATCCCAACACTGTATTGA	1313
QY	1021	tgggtgaaagacactttaaoccaagatgaagacacttgggtccaccccttaaggttgcag	1080
Db	1314	TGGGTAAAGAACTTTATAGCCAGATGAACACTGTGGGCCACCTTACAGCTTACAGAGG	1373
QY	1081	tgaatgcctgtgcttgttcccaacaccccaagtaagacatctcagaatagaactctat	1140
Db	1374	TGATGCTGTGGCTGTGTTCCCAACACCCCAATGATGACATCTCAGACATGACTTCTAT	1433
QY	1141	gccaggtcgtgtcgaatggcagggatcatgtagggagacatcgataaagggttccctatgtc	1200
Db	1434	GCCAGCGTGTCAGATGTGGGACGAGGTCAATGAGGAGACATGAATGAAGGTCTCTTATGCT	1493
QY	1201	ccctgcctctgtaatccaacgaaggggcctatctgtgttataggggcctgggaacttgatgg	1260
Db	1494	CCCTGCTCTGGAATTCACACGACCGGCTATGTGGGTTATAGGGGCTGTGGGACTTGAAATGG	1553
QY	1261	atgtctcaaaacccaacactgttggccaacaaglttgacccaagaatagatgataatgtc	1320
Db	1554	ATGCTTCAAAACCATATCACCTGTTGGCCACAAAGTTTGACCCCAAGATGATGATATGCT	1613
QY	1321	cttaagtgcttagaagaataactacgttataaaggcaactataaggacactgaactttgaac	1380
Db	1614	CTTCAGTCTTGAAGAAATCTTACATGTTAATGAAGCCATTTATGGACATGAACTTTGAGAC	1673
QY	1381	aacactatgaagcatttgcctacccgttgggcaagagcatgtacaacaatgctcaagaactg	1440
Db	1674	ACACTATGAGAGCGTGTCTACTACGTGTGGGGCAAGAGCATGTACAAACATGCTCAGAACTTG	1733
QY	1441	ctgggaacagtgctgggtgggaagccaaagggtcttgcaatctgtgtgaactccttagataaga	1500
Db	1734	CTGGGACAGTGTGTGGTGGGAGACCAAGGCTTTTGCAATTTGCTGGCACTCTTATAGGATAGA	1793
QY	1501	gggcgtgcattagaatttgggtlaagatagatccttgcctctgaaaatgtcctccgtgggtga	1560
Db	1794	GGGCGCTATTAGATGTGTGGGTAGTAGATCTTTTGCCCTTGGAAATGCTGTGGGTGA	1853
QY	1561	atgcgtctgtcttccataccccaaccctagtagtcttcccaactaacttctcactaagt	1620

Db 1854 ATGCTGCTGTTCTCTCAACCCCTTAACCTAGTACTTCTCCACTAATCTTCTCACTAAGT 1913

Qy 1621 gagaatgaactctgtgataagaagaagatgaagaagagatactgtgataagcacttgat 1680

Db 1914 GAGATATGAACTGCTGTGATAGGAGAGTGAAGAGGAGATGTGTGATAGACACTTAT 1973

Qy 1681 ttcaattgaatgcctgtgtaagctttccattctgtgagctgcctgtccataaattc 1740

Db 1974 TTCAATGTAATGCTGTGTAGCTTTTCCATTTCTGTGAGCTGCCGTTCTCAATATTC 2033

Qy 1741 caggttgatgctgtgagaggaactttgatgtaagaagaagacccctctgtactgt 1800

Db 2034 CAGGTTTGTAACGTGAGGAGAACTTGATGAAAGAGAAC--TTCTCTGTACTGTT 2091

Qy 1801 aacttaaaataaata 1816

Db 2092 AACTTAATAATAATA 2107

RESULT 7

AF102542 2128 bp mRNA linear PRI 31-JAN-1999

LOCUS AF102542

DEFINITION Homo sapiens beta-1, 6-N-acetylglucosaminyltransferase mRNA.

ACCESSION AF102542

VERSION AF102542.1 GI:4204683

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 2128)

AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.

TITLE Molecular cloning and expression of a novel beta-1, 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I branches

JOURNAL J. Biol. Chem. 274 (5), 3215-3221 (1999)

MEDLINE 99115671

REFERENCE 2 (bases 1 to 2128)

AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.

TITLE Direct Submission

JOURNAL Submitted (28-OCT-1998) Glycobiology Program, The Burnham Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037, USA

FEATURES

source

1..2128

/organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue_type="brain"

/dev_stage="fetus"

354..1670

/note="glycosyltransferase: core 2 Gnt-M"

/codon_start=1

/product="beta-1, 6-N-acetylglucosaminyltransferase"

/protein_id="AADI0824.1"

/db_xref="GI:4204684"

/translation="MVQMKRLCOLHYLMALGCMYMLLTVALLKLSFRLKCDSDHLGLES RESGSOYCRNLVNFLLKPAKRSINGGVWGDQEAIVLQALINLEKRLKKEPTDTH YLSLRDCEHFKARFIQPIKSEKEVEFPAYSMVHEKRENERLLRVAAPONTY CVNHYDEKSPERFKAVKAIISCPNVFIASKLVVYVAVSMRVDADLMCDLLOSY PWKFLNTGCTDFPKSNAEVQALKMLNGNSMESVPRKHKTAKYHFEVVDL HLNKKKRDPPYNTLMTGNAYIVASRDFVOHVLKNRKSQDLIMVVDVTSDEHLMV TLDARMMGSPVNHPRKDYIDSDMTSIRLVKMOGHEDIDGAGAYAPCSGIHQRAICV YGAGDLMMNLNHLANKRPDYDNLALOCLELYLKKAIYGRLEL."

BASE COUNT 569 a 477 c 526 g 556 t

ORIGIN

Query Match 95.5%; Score 1798.8; DB #: Length 2128;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1812; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 gatgtgctctctcaccctctctgtctcggtctccacctgtctccattctgtgacg 60

Db 294 GATTGTGTCCTCCCTCAACCTTCCTGCTGCTGCTCCACCTCTCTCTCCACTTCTGTGACG 353

Qy 61 atgtgtcaatgaagaagctctgtccagctgcaacttactgttgagctctgtgactatg 120

Db 354 ATGTTCAATGGAAGACACTCTCCAGCTGCACTTACTTGTGGCTCTGTGGCTGCTATATG 413

Qy 121 ctgtctgcaactgttgctctgtaaatcttcttccagttgaaagtgtgactctgacactg 180

Db 414 CTGCTGGCCACTGTGCTGCTGAATCTTCTTCAAGTGTGAAGTGTGACTGTGACCACTTG 473

Qy 181 ggtctgagctccaggaactctcaagcagctactgttagaatalactcttataattctg 240

Db 474 GGTCTGGAGTCCAGGGAATCTCAAGCCAGTACTGTAGAAATATCTTATATTTCTCG 533

Qy 241 aaacttccagaagaagctctatcaactgttcaagggtccaccggaggagaccagaagca 300

Db 534 AAACCTCCAGAAAGAGGTCTATCACTGTTAGGGGTCACCCGAGGGGACCAAGAGCA 593

Qy 301 gtgttccagctatctgataaaccctgaggttcaagaaagcagagaccttccagaagc 360

Db 594 GTGCTTCAGGCTATCTGAATTAACCTGAGGTCAGAAAGAGCGAGACCTTTCACAGAC 653

Qy 361 acccaactctctccctcaagagactgtgagcacttcaaggtcgaaggaagtacata 420

Db 654 ACCCACTACTCTCCCTCACAGAGACTGTGACACTTCAAGGCTGAAGAGAACTTCATA 713

Qy 421 cagttccacactgagaagaagagtgtaggttccctattgtatctatctatgtgattcat 480

Db 714 CAGTCCCACTGAGCAAAAGAGAGGTGAGTCCCTATTCATCTATGATGATATAT 773

Qy 481 gagaagattgaaacttgaagaagctgtagctgtgtagtccctccagaacataac 540

Db 774 GAGAAGATTGAAACTTTGAAGAGCTACTGCGAGCTGTGTATGCCCTCAGAGACATATAC 833

Qy 541 tgtgtccatgtgagtagaaggtccccaagaacttcaagaagggcgtcaagaactatct 600

Db 834 TGTGTCCATGTGAGTAGAAGATCCCAAGAACTTTCAAAGAGCGCTCAAGCAATATAT 893

Qy 601 tctgtctcccaaatgttctatagccagtaagctgtgtcggtgttattagctccctcgg 660

Db 894 TCTTCTTCCCAATGTCTTCTATACCAAGTAACTGTGCTTCCGCTGTGTTATGCTCTCTG 953

Qy 661 tccaaggtgcaagctgcaactcaactgcaatgcaagaagactgttccagaagctcagtcg 720

Db 954 TCCAGAGTGAAGCTGACCTCAACTGATGAAGAACTTGTCTCAAGCTCAAGTGCCTGAG 1013

Qy 721 aaatactctcgaatacatgtgtgagcaggaacttctataagaagcaatgtcagaagtgc 780

Db 1014 AAATACTTCTCCAAATACATGTGGAGGACTTCTTATAAAGCAATGCAAGATGTGCT 1073

Qy 781 caggtctccaagaatgtgaatgaggaataagcagaggtcagaggtaccctctaagcac 840

Db 1074 CAGGCTTCCAGATTTGAATGGAGAAATAGCATGGAGTCAAGAGTACTCTCTAAGCAC 1133

Qy 841 aaagaacccgcttggaatalactacttggagtagtgagagacacataacacaaacac 900

Db 1134 AAAGAACCCTCGTGAATATACCTTGAAGTAGAGAGACACATTTACACTTAACAC 1193

Qy 901 aagaagaagatctctccctctataacttaactatggttaagggaaatgtcactgtg 960

Db 1194 AAGAAAGAGAGATCTCCCTTATATTAATTACTATGTACAGGAAATGCCATTTGTG 1253

Qy 961 gcttcccaagatttgcctcaacatgttttgaagaacccctaactccaactgtttaa 1020

Db 1254 GCTTCCGAGATTTCGTCAACATGTTTGAAGAACCTTAATCCCAACACATGATTA 1313

Qy 1021 tgggtaaagaacactatagcccaagatgaacacccctgtgccaacctcgaagcgttgcag 1080

Db 1314 TGGGTAAAGACACTTATAGCCCAAGATGAACACCTGTGGCCACCTTCAGGTCACAGG 1373

Qy 1081 tggatgctgtgctgttcccaacaccccaagtagaactctcagaatgactatctat 1140

Db 1374 TGGATGCTGTGCTGTCCCAACCCCAAGTACGACATCTCAGACATGACTTCTATTT 1433


```

|||||
Db 781 AAAAAGCCCTGGAATATCACTTGTAGTAGAGAGACATTAACCTAACAC 840
QY 901 aagaagaagctctcccccataaactaactatgtttacaggaatgcgtacattg 960
Db 841 AAGAAGAAGAGTCTCCCTTATTAATTAAGTATTAAGGAAAGCTACATTGTG 900
QY 961 gctccagagatctcgtccaaactatgtttgaagaacccataaccacaactatgaa 1020
Db 901 GCTCCAGAGATTGCTCCAAATGTTTGAAGAACCTTAATCCCAACATGATTAA 960
QY 1021 tgggtlaaagaactatagaccagatgaacacactctggccacccctcaagctgacg 1080
Db 961 TGGTAAAGAGACTTATAGCCCAAGATGAACACCTGGCCACCTTCACGTCACGG 1020
QY 1081 tgaatgcctgctctgtttcccaacaccccaagtaacacatctcagaatgactctatt 1140
Db 1021 TGAATGCTGCTGCTGTCCACACCCCAAGTACGACATCTCAGACATGACTTCTATT 1080
QY 1141 gccagctgtgcaagtgcaggtgcatgagagagacatcgatgaaggtgctccattgct 1200
Db 1081 GCCAGGCTGGTCAAGTGGCAGGATCATAGGAGACATCGATTAAGGTGCTCCTTATGCT 1140
QY 1201 cctgtgcttgaaaccacagcgagctatctgcgtttatggggctggggacttgaaatg 1260
Db 1141 CCTGTGCTGGAATCCACACGCGGCTATCTGCGTTATGGGGCTGGGGACTTGAAATGG 1200
QY 1261 atgtctcaaaaccatcaccctgttggccaacaagtgttgacccaagatgataatgct 1320
Db 1201 ATGCTTCAAAACCATCACCTTGTGGCCAAAGTTTGACCAAAAGCTGATGATATGCT 1260
QY 1321 ctcaagtcttagaagaatactactacgtatataagggacatcttggaacttgaa 1377
Db 1261 CTTGAGTCTTAGAAGATTAAGTATTAAGGACATCTAATGAGACTGAACTTTGA 1321

RESULT 9
AF038650 1317 bp mRNA linear PRI 25-MAR-1999
LOCUS AF038650 Homo sapiens core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase
DEFINITION (C2/4GnT) mRNA, complete cds.
ACCESSION AF038650
VERSION AF038650.1 GI:4511880
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eulheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1317)
AUTHORS Schwiientek,T., Nomoto,M., Levery,S.B., Merx,G., van Kessel,A.G.,
Bennett,E.P., Hollingsworth,M.A. and Clausen,H.
TITLE Control of O-glycan branch formation. Molecular cloning of human
cDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase
forming core 2 and core 4
JOURNAL J. Biol. Chem. 274 (8), 4504-4512 (1999)
MEDLINE 99143102
REFERENCE 2 (bases 1 to 1317)
AUTHORS Schwiientek,T. and Clausen,H.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-1997) Dental School, Faculty of Health Sciences,
Copenhagen University, Noerre Alle 20, Copenhagen 2200, Denmark
FEATURES
source
1..1317
location/Qualifiers
1..1317
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1317
/gene="C2/4GnT"
1..1317
/gene="C2/4GnT"
/note="core 2/4-GnT"
/codon_start=1
/product="core 2/core 4
beta-1,6-N-acetylglucosaminyltransferase"
CDS

```

```

/protein_id="AAD21525.1"
/db_xref="GI:4511881"
/translation="MNQKRLCOAHYMLALGCVMLATVALKLSFRKCDSDHGLRES
RESOSYCRNLIYFLKLPARKRSINCSTVRGDEAVLQALNNLEPKRKREPTDTH
YLSLTFRCEHEKARFKIQLPSEEEFPLAVSMVHEIKENERLRAVAPONTY
CVHDEKSPETFEKAVRAIISCEPNEVIAKLVAVYASRVOADNCMEDLQSSV
PMKRYFLNTGCTDEPIKSNAEVQALKMLNRSNMESEVPKHEKRYHEEVDRDL
HLNFKKQDPPEYNTMETGNAYIVASDFVOHYLVKNKSOOLIEWAKDYSPDEHLMA
TLORARMPGSPVPHPKYDIDSMYSIRLVKMOGHEDIDKAPAYACSGIHORALICV
YGAQDLWMQNLQNHILLANKRDPKDYDNALQCLEYELKRAIKYTEL"
BASE COUNT 359 a 307 c 322 g 329 t
ORIGIN
Query Match 69.9%; Score 1317; DB 9; Length 1317;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 61 atgtlcaatgaaagagactctgccaactgcatlactctgttggtcctgtcgtatag 120
Db 1 ATGCTTCAATGGAAGAGACTCTGCCAGCTGCATTACTGTGGCTGTGGCTGCTATATG 60
QY 121 ctgctggccactgtggtcctgaacattccttcaagttgaagtgtgacttgaccattg 180
Db 61 CTGCTGGCCACTGTGGCTCTGAACCTTCTTCAAGTTGAAGTGAAGTGAAGTGAAGTGA 120
QY 181 ggtctgagtcgaaggaatctcaagccagtaactgtaagaaatctgtataattcctg 240
Db 121 GGTCTGAGTCCAGGGAATCTCAAGCCAGTACTGAGAAATCTTATATATTTCTG 180
QY 241 aaactlccagaaagaagtlcatacaactgtlcaaggggtccaccggaggagca 300
Db 181 AAACCTCCAGCAAAAGAGTCTATCAACTGTTCAGGGGTACACCCGAGGAGCAAGAGCA 240
QY 301 gtgtctcaggtactctgaataacactgaggtgcaagaagcgagaccccttccagac 360
Db 241 GTGCTTCAGGCTATTTCTGAATAAACCCTGAGGTCAGAGAAAGACCGACCTTTTCACAG 300
QY 361 acccaactacccctccctaccagagactgtgagacacttcaagggctgaagaagatcata 420
Db 301 ACCCACTACCTCTCCCTTACCAGAGACTGTGAGCACTTCMAAGCTGAAAGAAAGTTGATA 360
QY 421 cagttcccactgagcaagaagaggttggagttccctattgtcactctatagttgattcat 480
Db 421 CAGTTCACAGTGAAGCAAGAGAGTGGAGTCCCTATTTGATCTATGATGATGATTCAT 420
QY 481 gagaagattgaaacacttggaaagctactgagactgtgtatgctccctccagaacataac 540
Db 481 GAGAAAGATTGAAACCTTTGAAAGGCTACTGCCAGCTGTGTATGCCCTCAGAACATATATAC 480
QY 541 tgtgtccatgtgagatgagaaggtccccaagaacttcaagaagggggtcaaaagaactatatt 600
Db 541 TGTGTCCATGTGAGATGAAGAGTCCCAAGAACTTCAAGAGAGGCGTCAAGCAATTTATT 540
QY 601 tcttgctcccaaatgtcttccatagaccagtaagctgtgtcgggtttatagctccctcgg 660
Db 601 TCTTGCTTCCCAAAATGCTTTCATACCAAGTAACTGTGCTGGGTGTTATGCTCCCTGCG 600
QY 661 tccagagtgcaagctgaccctcaactgcatlgaagaactgtgtccagaagctcagtcgctgg 720
Db 661 TCCAGAGTGCAAGCTGACCTCAACTGCAATGCAAGCTGTCTCCCAAGCTCAAGTGCCTGG 660
QY 721 aaatactcccgaaatagatgtgagacgagacttccatataagaagagaagcagaagatgtc 780
Db 721 AAATATCTTCTGAAATACATGTGGAGGAGACTTTCTTAAAGAGCAATGACAGATGGTTC 720
QY 781 caggtctcagaagatgtlgaatgaggaataagcatgagtcagaggtlcaactcctcctaagcac 840
Db 781 CAGGCTTCAAGATGTTGAATGGAGGAATAGCATGAGATGACAGAGTACCTCTCAAGAC 780
QY 841 aaagaacccgctggaataatacactttagagtagtgagagacacattcaactaaccaac 900
Db 841 AAAAGAAACCCGCTGGAATATATCACTTTGAGGTAGAGAGACATTAACCTTAACCAAC 840

```

QY 901 aagaataaagatctccccccttaattactatgtttacagggaatgcgtacattgtg 960
|||||
Db 841 AAAAAAAGATCCTCCCTTATTAATTAATGATGTTACAGGAATGCGTACATTGTG 900
QY 961 gctcccgagattcgttccaacatgttttgaagaacctaatacccacaacatgttaa 1020
|||||
Db 901 GCTTCCCGAGATTTCGTCCACATGTTTTGAAGAACCTTAATCCCAACACTGATTGAA 960
QY 1021 tgggtaaagacactatagaccagaatgaacacctctgagccaccttcagcgtgacag 1080
|||||
Db 961 TGGGTAAAGACACTTATAGCCAGATGAACTCTGGGCCACCTTCAAGCGTGACGG 1020
QY 1081 tggatgctgctctgttcccaaccacccaagtaagacatccagacatgaacttatt 1140
|||||
Db 1021 TGGATGCTGCTCTGTTCTCCCAACACCCCAAGTACACATCTCAGACATGACTTATT 1080
QY 1141 gccaagctgtcaatgagcagggtcatgaggagaaatgataaggtgctcttaagct 1200
|||||
Db 1081 GCCAGGCTGGTCAAGTGGCAGGGCTCAGAGGAGACATCATTAAGGCTGCTCTTATGCT 1140
QY 1201 cctctgctctgaaatccaccagcggatctgctgattatgagctggagacttgaattg 1260
|||||
Db 1141 CCTGCTCTGGATTCACACGCGGGCTATCTGGGTTATGGGCTGGGACCTTGAATTGG 1200
QY 1261 atgcttcaaaaaccatcaacctgttggccaacaagtgtgacccaaaggtagaatgact 1320
|||||
Db 1201 ATGCTTCAAAAACCATCACCTGTTGGCCACACAACTTTGACCCAAAGATGATGATGCT 1260
QY 1321 ctccaatgctctagaagaataccctagcttataagggcatcatgaggtgaacttga 1377
|||||
Db 1261 CTTCAGTGCTTAGAGAAATACCTACGTTATTAAGGCCATCTATGGAGCTGAACCTTTGA 1317

RESULT 10
AF318573 108873 bp DNA linear VRL 18-MAY-2001
LOCUS AF318573
DEFINITION Bovine herpesvirus 4 long unique region, complete sequence.
ACCESSION AF318573 AF271211
VERSION AF318573.1 GI:12802528
KEYWORDS
SOURCE Bovine herpesvirus 4.
ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae.
REFERENCE 1 (bases 1 to 108873)
AUTHORS Zimmermann, W., Brodl, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and
Goltz, M.
TITLE Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
Identification of an Origin of DNA Replication
JOURNAL J. Virol. 75 (3), 1186-1194 (2001)
PUBMED 11152491
REFERENCE 2 (bases 1 to 108873)
AUTHORS Zimmermann, W., Brodl, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and
Goltz, M.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Project P24, Robert Koch-Institut, Nordufer
20, Berlin D-10555, Germany
FEATURES
source
1. 108873
/organism="Bovine herpesvirus 4"
/db_xref="taxon:10385"
1. 108873
/note="long unique region; EUR"
complement(445..633)
/note="ORF B01"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pB01"
/protein_id="AAK07921.1"
/db_xref="GI:12802529"
/translation="MEGDYSMAVFGKRTKRLDNLNSAGNEVCALVLRKNINVSAGC
CAVKRNNSFHSLPGMWP"

CDS
614..4480
/note="segment protein; ORF3 BORFA1"
/codon_start=1
/product="v-FGAM-synthase"
/protein_id="AAK07922.1"
/db_xref="GI:12802530"
/translation="MDTSPSIOLCFPVSRRAPLSNCOVIOOFTSGOLPLTSVLOY
EGVFNVLVGEBOAVREPLVPLDLKLFSSGELTPRESSILKTDSRLSFYGVPSA
PETALSRLDVALIQQSGSLAPMSNLTQLIKLYSLTISIKMLNPSKILHSTLRAL
CPDPSLFRPYKSEIYVGHAFHGAPOGLSPKDDLLCAFGNGLSAGSGEMFT
POHGPOTIVNOMCKTAYLVPDVYMTISTSTPLSGHMDPLGTYTGOLDOVARNC
LKPLNYSNOMGEKANEVSSOLGIASVTSITSPSSRYELPPTYDIDMAYINMRCD
SLPCHOGVPEPIVESHPLDVSPEGRVAGCSERNTTPTDLPRETKINITYOIGT
FTSHEDSGREYRCLGALKYLMSEFLRSQVKGISAVKVCETMAKLISVCGGTGIG
ITYOSALPEPLNGLLPEFASNRQOMTMKQHLNIAVAVLVALEAPDNTDTVLK
ACCAIHGCOYKVLGRYOKPHIYVKDKDEQWVRKISLRKSPSPYLSITKKLKOS
ENMODSEMDPLPSLYLLEILKHPAGCQDFIKYKIHDRSLRSVAQOOGIGARDI
PISDYSIIYSDLSIAAAPRGSENMATRESLDDIPLVNEPEPVGIGSIAIECPPLS
TYEPIKGAHVAVTESLNTVSNAPFKRIDVCTFPNTWQTTDSHFGITELMRAREP
CSQIGICVFTSCITSNRQKCLNLSLITLVATATACADVTOGLFPLDLPSSII
VWLPSTEXHEPFGTVISOLFSDPSGSHINIDPLVYKLLITVTKHLMHESIISCHDV
CGGLITACFEMAYAGASIALTVPODEDPVFLFSETPGLAVEPVAVSTVORHE
TSDIITYEDVGRPLPSVANTFTVYSKERLIFRESINEMVEMNRHSTKOMYFCEY
OGTPEDPDKDLOLHITFPOPSMGPHGPYKHYOVNVYILPCTNTPDLSILVALEAGRVN
LVSTFDKIVKIVYDTTNVFGICLIGATIDEATLGDKAISMTTHNSVYGEELKLI
NSPDVFSLIGHACQILFENKFMGYNKSNMTLCKENVSLMESRLNFEIPENTH
AVALOSMKGSLPBGWIOGTHLGFHAFENKSNMTLSTHGAVATQFGADISAPALYTP
QNPATGYTISGLCSADGRHALDLHPGSELSNNLMQMPHLPKMTPLKVSFWKRMFLDLH
IWANKREMDQPPPHDPPLRNITKVM"
4608..4811
/note="ORF B02; BORFA2"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pB02"
/protein_id="AAK07923.1"
/db_xref="GI:12802531"
/translation="MASPQYCHRWVRVYGPCKKSAKISGYVDROYTHQJTSKNAYS
GMSPOCPSCADLNKLFVSOCPLS"
5286..5513
/note="ORF B03"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pB03"
/protein_id="AAK07924.1"
/db_xref="GI:12802532"
/translation="MGFNPMVLVYFHCNCTPTSLKHGILIKYIKISYLIRATFLRVN
HICPKRDRPRPGCGPPSLIIVGLALAK"
5943..9347
/note="ORF 6"
/codon_start=1
/product="single-stranded DNA-binding protein MDPB"
/protein_id="AAK07925.1"
/db_xref="GI:12802533"
/translation="MAMKTAQGOQYEDNIGSKAPIGCGYIYVYKKEPPTEASLIG
NGFPGSVASVPLRGLCTVYERGPPIVYKVRKALIDTTPYSVTVYHRVIMESNDIC
ETPIRFGTGLPOLCOETRLDREFGTHSPSSNTINNISLDDICGTNCIDBRLGVIYIT
EGFKERLYHGMHPLIGHVEVOIISPOHAKRILPLDEDLFSKIAFRDKRHSADVDS
OYLFNSLYTAIAQAIRVDVAIVIOAMESQAFVADHKMKPVVKKRDPSSSSRGVGL
SLIMDSVASELAVSYGLSFIDAPDMSMLDLYTSMPLFVDCITTEEDRLALHANNLK
OSIHVNTOLFSTNSMLYLRIOKONPKTKGDLNLYNSYILOHGLSYSEAOADDEYGO
PVFGVOSNLSLGSSTYTIHHLAVASMCNTLARYCYLOFOHOKSONOYNSIANSY
VGTAASSDMCNLCQGGKPRVVCINTLPLRYLRDPRPIITSHRDPYITIGSVMYNDD
LLGNFASFEREKDEGAQVEEVQRYTWDTLRLKLEMGKIKDTTPPEKLVSDIPSF
IKVFDIDAHVDTEVLKFINCWANNININREKYSIHNYIOCVCAWVFOQPPCPVLOL
YKSKILTIIDVCLPNCMAYEODNSMCGPEMKIKMYOGLMTNFKGCCGPRGVLTG
AELKVMHSDPPCDFDFDVDAINGIFSPKTVOYIRAMLVVKTIKRRIIESNSG
SRAIOSGFLSKSSKKNKYIVYPRYKFLINTYIKIMFPOTKMSLPMHNFSSKKRDIPL
ITGVRDQDLVALANTIEYNSKLHSELDVLDIIPDLVITAKIKLNNALIRACGQTOFY
STTLCLVPTVTVGEEYVPHVLENEEELGTBOYLQVONRTAQVQVATLKGDVAMQ
KRLPITIVPMVINKYTGINGNGIFHCQGLGFEMKRGVDRNLIFENAPKFKOSTAAM
RKHVEMPIVDNLIKRTYVPSSETEENISIRLOSLIDLEEDONONIPADVYLELVK
LGOACADPLTADLOFYIGEVYIMSDIELISRLQSIDACVAPWCEDESSVSLGSEVYSEE
QLEFTIGLEQOTTSAFTSEDFQASGLSTIIAAGKKRLNCLMSDFDL"

OY	763	ggcttcagatggttgaatgaggagaataagcattggaatgaaggtaactccctaagaacaa	842
Db	108044	GGCCCTCAAGATGTTGAAGGGTAAGACAGTGTGAGTGTTGAGTRCCTTTGATGCCAA	108103
OY	843	agaaaaccgcgtggaataatactattgagttagtgagacaatatcaactaactaacaaa	902
Db	108104	AAAAAATGCCTGGAATAATCCGTTATGAG--GTGACAGCACACTGTACCCTACCGACAA	108160
OY	903	gaagaagagatcctccccctataatttaacttgttttaaaggaatagcttacattgtgc	962
Db	108161	GATGAAGAAGACCTCCCCCTGATTAATTTCACCATGTTCCACAGGGAATCCCTATTTGTGGC	108220
OY	963	tcccgaagattcgtgccaaacatglttligaagaacctaatcccaactgatitgaag	1022
Db	108221	CCTCTGAGCGCTTGTGTCCAACATGTTCTTACACAACCTTAATCCCAAAGACTGTTGAATG	108280
OY	1023	ggtaaaagaacattatagccagatgaacaactcttggscaccttaagcgltgacgltg	1082
Db	108281	GGTCAAAAGACACCTTATAGCCCCGAGGAACACCTCTGGGGCCACCTTCACGTCGCTCGTG	108340
OY	1083	gatgctgagctctgtttcccaaccacccccaagaagaacatctaagacatfacttatgc	1142
Db	108341	GATGCCCTGCTTGTTCCTCAGCCACCCCACAGATACATCTCAGACATGACATGCCATGCC	108400
OY	1143	cagagctgtgtcaagtgtgcaaggtcatgataaggagaacalcga laaaggltgtccttatgtcc	1202
Db	108401	CAGGCTGCGTCAGTAGTGACATACACAGAGGAATGTCACGATGTGGGGCCCTTATGCACC	108460
OY	1203	ctgctctgtgaatccacacagcgggctatctgcgctttatgggctgtgggaacttaattgat	1262
Db	108461	CTGCTCTGGAATCCATCGGAGGGCCATCTGCATTTTACGGGGCCGGGACCTTACTTGAGAT	108520
OY	1263	gcttcaaaaccatcacctctgttgccaacaagtlttgaccaagaqitaagatgataatgctct	1322
Db	108521	TCTCCAGAACCATACCTCTTGGAACAACAAGTTTACCCGAGGGTGTGATTAACCTCCT	108580
OY	1323	taagtgtcataagaataactactacgttatataaggccatctatgygaacttgaactltgagac	1382
Db	108581	GCAGTCTTTAGAAAGATGATACCTTCGTCATTAAGGCCATCTTGTGGACATCTTGTGACCTCG	108640
OY	1383	actatgagagcgcttgctactctgttgaggcaagaagcatgatacaaatgctcagaacttgtc	1442
Db	108641	ACTGT-----GGGCCAGGAGACTGTCTCCCAACATGCCACGAGGCGCTACT	108681
OY	1443	gggacacagtgt-ggggtgggagaccagggcttgcnaattgtgtgcatcctttaagataaag	1501
Db	108682	GGTACTGTGTGGGGGGGAGAGAGGGCGTGTGAATTCATGTGCAATCCCAAGGTCAGAG	108741
OY	1502	ggctgtctatlagattgttgtaagtag	1528
Db	108742	GGCTGCTTTGATGGGCTCAGAGATGTG	108768
RESULT 11			
LOCUS	AF231105	2017 bp DNA linear	VRL-26-MAY-2000
DEFINITION	Bovine herpesvirus 4 beta-1,6-N-acetylglucosaminyltransferase (BOFV-4) gene, complete cds.		
ACCESSION	AF231105		
VERSION	AF231105.1	GI:8096688	
KEYWORDS			
SOURCE			
ORGANISM	Bovine herpesvirus 4. Bovine herpesvirus 4 Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae.		
REFERENCE	1 (bases 1 to 2017) Vanderplassen A., Marikne-Gorlaynoff N., Lomonte P., Suzuki M., Hiraoka N., Yeh J.C., Bureau F., Willems L., Thiry E., Fukuda M., and Pastoret P.P. A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded by bovine herpesvirus type 4		
AUTHORS	Proc. Natl. Acad. Sci. U.S.A. 97 (11), 5756-5761 (2000)		
JOURNAL			

MEDLINE	20283898
REFERENCE	2 (bases 1 to 2017)
AUTHORS	Vanderplasmchen J.A., Markine-Gorjanyoff,N., Lomonte,P., Suzuki,M., Hiraoka,N., Yeh,J.-C., Bureau,F., Willems,L., Thiry,E., Fukuda,M. and Pastoret,P.-P.
TITLE	Direct Submission
JOURNAL	Submitted (03-FEB-2000) Immunology-Vaccinology, University of Liege, Faculty of Veterinary Medicine B43bis, liege B-4000, Belgium
FEATURES	Location/Qualifiers
SOURCE	1..2017
	/organism="Bovine herpesvirus 4"
	/strain="V. test"
gene	/db_xref="taxon:10385"
	227..1549
CDS	/gene="BOREF3-4"
	227..1549
	/gene="BOREF3-4"
	/function="Forms core 2, core 4, and I branches"
	/note="last ORF at the right end of the L-DNA; similar to Homo sapiens core 2
	beta-1,6-N-acetylglucosaminyltransferase, mucin type; pBOREF3-4"
	/codon_start=1
	/product="beta 1,6-N-acetylglucosaminyltransferase"
	/protein_id="AAF72001.1"
	/db_xref="gi:8096689"
	/translation="MKMGKKKIKLPCGHNLALGCMYLLAVVSLRLSRFKCDVDSLLESDPOSCHCRDMLYNLSLKLPARKISNCSITGGDDAAVVQALLDNLEVKKRPPLTDYILNTRODCERPKAOKFTIOPLSKEELDFPIASLYNVVHEKIENFRILRAYAPDNIYCVDHYKSPERRFEAKVKAIISCPRVPMAKSITLVPAVSMSRNOADINLCMDLIQS SVSKYLINTCGTDFPIKTAEWLAIAKLKRGKSMSESVSSKKRKRYRTREVTDTLPYSKITDPPNPDLPEPTGNAYEVASRAVQVLNDPKSQILLVEWKDYSPDEH WATLORAPMPGPVSPSHKVAHIDMTALAARLVKQVHGSDVMGAPAPACSGIHRRAD CTVAGDLYWLTIOMHHLLANFDRVDVNQLCELEYLRKAIIYTGL"
BASE COUNT	493 a 504 c 535 g 485 t
ORIGIN	
Query Match	53.2%; Score 1001.4; DB 14; Length 2017;
Best Local Similarity	80.6%; Pred. No. 1.6e-308;
Matches 1230:	Conservative 0; Mismatches 271; Indels 26; Gaps 4;
Dn	6 tgcctcccccacacctccttcgtgctgcggagtctcacactctgccattctcgtagcatgt 65
Dd	178 TCTCTCTCTCCACCCTTGCTGCTGTGCCAACACTCTCTGATCTTCATCCCATGAATGCC 237
Dy	66 tcaatyggaagaga---ctctgcacgctgcatacttcttgtygctctcggctgataatgct 122
Dz	238 TGGGTGGAAGAAGAGACTTTGCCCGGAGCATCATCTGTMGGGCCCTGGGCTGCTATATGCT 297
Dq	123 gtctggccaactgtggtctctgaaactcttcttcaggttgaagtgtgacttgaccacttgg 182
Dv	298 GCTGGCCGCTTGTTCTCTGAGACTTTCCTCCAGGTTTTAAATGTGANTGATTCCTTGGAA 357
Dw	183 tctggagtcacaaggaatctcaaacgacagtaactgtgaaatatcttgtlaaatlctcgaa 242
Dx	358 TCTGGAGTCTCAGGAGACTTCCACAGTCAGCACTGTAGGAGCAATGTTGTACAATAAGCTGGA 417
Dy	243 acttccagcaagaagtgctatcaactgtltcaagggttcaaccogaggggaccaagaagcagt 302
Dz	418 GCTGGCAGCAAGAAGATCCATCATCTGTCTTGCGATTACGCGCAGGGGAGCACAGGAAGCAGT 477
Dq	303 gctttaaggtattctgataaactcttgagggtcaaagaagaagsgaagagcccttcacaagac 362
Dv	478 GGTCAGAGGCCTCTCTGAGACAACTTGGAAATGAAGAAAGACGGCCGCTCTCTACTGACAC 537
Dw	363 caactactctctccaccacagagactgtgtagaacttcaagsgctgaaaggaagtacataca 422
Dx	538 CTATTACGTCAACATTAACACAGGAGCTGTGAGCGCTTCAAAGGCCCAAGGAATGTCATGACA 597
Dy	423 gtcccccaactggacaagaagaaggtgagtgctccatattgatatacttatcgtgatacga 482
Dz	598 GTTCCCACTGATGAAGAAGAGACTGTACCTCCCATTGTGCTACTGATGAGTGCTCATGA 657

QY	483	gaagattgaaacacttgaagaagcctacgcgagcctggtatgatgccctcagaacatactc	542
Db	658	GAACATTGTGAACCTTTTAACCGCGCTGCTGCAGCGCGTATATGCCCCCTTAGAACATATATCTG	717
QY	543	tgtcatatgtatgaagaagtcctccaganaactcttcaagaagcgctcaagaacattcttc	602
Db	718	TGTCACAGTGGATGTGAAGTCCCGAGAGACTTTCAAGAGGCGGTCAAGGCCATATATTC	777
QY	603	ttgcttcccaaatgltcttcataagcagtaagctggttgcggttggttatgcttcctctgctc	662
Db	778	CTGCTTTCCCCCAATGCTCTTCATAGCGCACTAAGATTGGTCCCGGTGTTATGCTCTCGTGTC	837
QY	663	cagggttgaagctgacctccaactcagatgaggaagactcttcagagctcagatgscgtggaa	722
Db	838	CAGAGTGCNAACCTACCTGAACTGTATGGAAGACTTCTCCAGAGCTCAATGTCATATGAA	897
QY	723	atacttcctgaatacatatgltgagcgagacttctcctataaagaagcaatgcaagatgttcca	782
Db	898	GTACTTACTGTAATCATAGCGGAGACGACTTCCCCATTAAGACCAATGCCAGATGGCTCT	957
QY	783	ggcctctaaagtggttgaatgagggaatagcatatggaatcgaagttactccttaagcaaa	842
Db	958	GGCCCTCAAGATGTTGAAGGTTAAGAACAGATGAGATCTGAGGTACTCTTCAAGTCCAA	1011
QY	843	agaaacccgcctggaataatacactcttgagtgtagtgaagacacataacctaacaacaa	902
Db	1018	AAAAAATGGCTGGAAATATACGCTGTAG--GTCACAGACACACTGTACCTACACGAA	107
QY	903	gaagaaggaatccctcccccataaattaaactatglttcaaggaatgctgatactgtgctc	962
Db	1075	GATAAAGACCCCTCCCTCGTAAATTTATACCATTGTTACAGGGGAATGCCATATTTGTGCG	1133
QY	963	ttcccgaaatttgctccaacatggtttggaagaaccctaaatccaaactgattgaatg	102
Db	1135	CTCTCGAGCCTTGTCCAACTATGCTTTAGACACACCTTAATATCCAAATCTGTTGATG	119
QY	1023	ggtlaaagaacatlatagccacagaatgaacacctcttgggccaacctcaagcgtgacagtg	108
Db	1195	GGTCAAGACACATATATGACCCTCCGACGAACACCTGTGGCCACCTTGAGCTCGTCGCTG	125
QY	1083	gatgctctggtctctgttcccaacaccccagaatgacacatctcagaatgactctattgc	114
Db	1255	GATCCCTGGCTCTTCTTCCATCCACCCAAAGTATACATCTCATGACATCGCATCGC	131
QY	1143	caggctgtgtaagatggcagaggtcatatggaggagacatatgataaggtgtctcttatgctc	120
Db	1315	CAGGCTCGTCAAGTGGCAGTATCCACGAGGGAGATGATGAGATGGGGCGCTTATGCAAC	137
QY	1203	ctgctctggaatccacacagcgggctatctcgttataatgagctctgaggacttgatgatt	126
Db	1375	CTGCTCTGGATTCATCGAGAGGCGCATCTGCATTTAGGGGGCGGGACCTGTACTGAT	143
QY	1263	gcttcaaaaacataactctgttggccaacaagtttgaaccaaaagtatgataatgctct	132
Db	1435	TCTCCAGAACCATATACCTCTTTGGAAACAAGTTTGACCCAGGCTGGATATTAACGCTCT	149
QY	1323	tcaagtgtcttagaagaatacctacgtctataaagccatctatcggagactaaacttgaagac	138
Db	1495	GCAGTGGTTAAGAGATGATACGTACTAATTAAGGCCATCTATGGGACATGAATTTGA	154
QY	1383	actatgagagcgttgtctactctgtggtggcaagaagcatgtatacaacatgctcagaactgtc	144
Db	1550	-----GCTGCAACGGGTGGGCGCAGGGCACTGTGCCAACATGCCAGGGGCTACT	159
QY	1443	ggagcagatgt-gggtggggaacacagggcttgcaactctcgtgacactctttagataaag	150
Db	1596	GGTACTGTGTGGGTGGGGGACAGAGGCTGTGGAATTCATAGGCATCCCCAGGGGTCAAG	165
QY	1502	ggctgtcatatagattgtyggtgaagtag	1528
Db	1656	GGCTGCTTCGATGGGTGACAGATTTTG	1682

RESULT	12	4179 bp	DNA	linear	VRL 06-JUL-1995
BHV4REFG	BHV4REFG				
LOCUS	Bovine herpesvirus type 4 DNA for nonconserved region F (DN599 like strain)				
DEFINITION					
ACCESSION	246386.1	GI:562278			
VERSION					
KEYWORDS					
SOURCE	Bovine herpesvirus 4				
ORGANISM	Bovine herpesvirus 4				
REFERENCE	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae.				
AUTHORS	1 (bases 1 to 4179)				
TITLE	Lomonte, P.				
JOURNAL	Direct Submission				
MEDLINE	Submitted (20-OCT-1994) Patrick P.L. Lomonte, Virology, University of Liege, Faculty of Veterinary Medicine, Bd de Colonster, Sart Tilman, Liege, 4000, BELGIUM				
FEATURES	2 (bases 1 to 4179)				
SOURCE	Lomonte, P., Bublolt, M., van Santen, V., Keil, G.M., Pastoret, P.P. and Thiry, E.				
FEATURES	Analysis of bovine herpesvirus 4 genomic regions located outside the conserved gammaherpesvirus gene blocks				
FEATURES	J. Gen. Virol. 76 (Pt 7), 1835-1841 (1995)				
FEATURES	97201604				
FEATURES	Location/Qualifiers				
FEATURES	1..4179				
FEATURES	/organism="Bovine herpesvirus 4"				
FEATURES	/db_xref="taxon:10385"				
FEATURES	/clone="EcoRI H (Bublolt et al., 1990) tot"				
BASE COUNT	1192 a 927 c 951 g 1109 t				
ORIGIN					
Query Match	51.0%; Score 960; DB 14; Length 4179;				
Best Local Similarity	80.9%; Pred. No. 3.3e-295;				
Matches 1180; Conservative	0; Mismatches 245; Indels 33; Gaps 4;				
OY	72 gaagagactctgccagctgcattactctgttggtctcttggtctgctatagctgctggccac	131			
Db	2649 GAAGAGCTTTGCCGGGAGCATCATCTGTGGCCCTGTATGCTGTCGGCGT	2708			
OY	132 tctgctctgaactcttcctcaagttgaagtgtaactctgacacctgggtctgagtc	191			
Db	2709 TGTTCCTGTGACTTTCCCTCAGGTTTAAATGTAATGATTTCTTGATCTGGAATC	2768			
OY	192 cagggaatctcaaaagccagctactgtgaagatactctgataattcttcgaaactctcagc	251			
Db	2769 CAGGACTTCCAGAGTCACTGTGAGGACATGTTTAAATAGCTGAGAGCTCCAGC	2828			
OY	252 aaaaagctctatcaactcttcagggtgtacccgaaagggaaagagagctgagtc	311			
Db	2829 AAAGAGATCCATCAACTTTGTGGATGACCGCAGAGGACCAAGAGAGTGTCCAGGC	2888			
OY	312 tatctgataaaccctggaggttcaagaagaagagagctcttcacagacaccacact	371			
Db	2889 CCTCTTAACCAACCTGGAGAGTGAAGAAAAAGCGTCCCTCTCACTGGCACCATTACT	2948			
OY	372 ctccctccacagagactgtgagacttcaaggtctgaagaaagtcatlacagttccact	431			
Db	2949 CAACATATACAGAGACTGTGAGCGCTTCAAGGCCCAAGAGAGTTCATACAGTCCACT	3008			
OY	432 gagcaaaaagaagtgagagcttccctatctgcatactctctatgtgatcatatgaaatga	491			
Db	3009 GAGTAAAGAAAGCTGACTTCCCATTTGCTTACTCGATGATGTGGTGAAGAGATTGA	3068			
OY	492 aaactttaaagagctactgcagactgtgtatgcccctcaagacalatactgtfctcatgt	551			
Db	3069 AAACTTTAAAGGCGTCTGCGAGCGGTATTATGCCCTCAGAAACATATATCTGTGTCAGCT	3128			
OY	552 ggaatgagaagctccacagaacttccaagaggcggtcaagaacatattctctgtctcc	611			

Db	3129	GGATGTGAAGTCCCGACAGACCTTTCAAAAGGAGGGGGCAAGGCCATTATTTCTCGTTCC	3189
Qy	612	aaatgcttctcaatagccagtaagctgtgttcgggtgtgttatatgcttcctcgtgtccagggtgtga	671
Db	3189	CAATGTTCTTCATGGCCAGTAAAGTTGGTCCGGGGTGTATATGCTCTCTGTCACAGAGTGG	3248
Qy	672	agctgacctcaactgcattgaagaactgtgtccagagctcagtcgctgtggaatacctcct	731
Db	3249	AG-----CTGTTTGGAAACACTTGTGCTCCAGAGCTATAGTCCATGGAAGTACTTACT	3298
Qy	732	gaatacatgttggagcgcagcttctctctaaagagcaatgcagaagatgtgtccagctctcaa	791
Db	3299	GAATACATGCGGGGACACACTTCCCTCCATTAAGAACCAATGCCGAGATGTCCTGGCCCTCA	3358
Qy	792	gagtgtgaatgggagagatgacatgcatgtatgcagaagttactcctctaagcaacaagaacccg	851
Db	3359	GATGTGTGAAGGTAGAGACGTTATGGAAGTCTGGATTCCTTTGTAGTCCAAAAAATAATCG	3418
Qy	852	ctggaatatcatcttbaagttagttgagagacatatcaacctcaaccacaagaagaaga	911
Db	3419	CTGGAAATTAACCGCATATGAG---GTGACAGACACACTGTATCCCTACCACAGCAAGATGAAGA	3475
Qy	912	tctctcccttataathtaactaactatgtttacagggaaatgcataatgtgtgtctccgaga	971
Db	3476	CCCTCCCCCTGATATTAATTACCCTCATGTTTCACAGGAATGCTATTTTGTGGCCTCGAGC	3535
Qy	972	tttgcgtcaacatggttttgaagaaccttaatacccaacaactgattgaatgtgtataaga	1031
Db	3536	CTTGTGTCAACATGTCTTTAGACACACCTTAATTCCTCAAAACACTGTTGTAATGGGTCAAGA	3595
Qy	1032	caactatagcccaagatgaagaacctcctgggcccaccttcagcgtgcaaggttgaatgcctg	1091
Db	3596	CACCTATAGCCCGAGSAGAACCTCTGGGGCACCCCTTACAGCTGCTCGCTGGATGCTCGG	3655
Qy	1092	ctctgttcccaaccaaccaagtagacacatctcagaacatgactctatattgcgaagctgt	1151
Db	3656	CTCTGTTCTTCATAGCCACCCCAAGTATCACTATCAGACATGATGCGCATGCGCAGGCTCGT	3715
Qy	1152	caagtggcaggggtcatatgagagagacatgataagagtgctctcttctatgtctcctctg	1211
Db	3716	CMAGTGCAGTACCAACGAGGAGGAATATCAGCATGAGGGGCGCTTATGACACCTGTCTGG	3775
Qy	1212	aatccaccagcgggtctatctgcgttttatatggggctgggagcttgaatttgaatgtctcaaa	1271
Db	3776	AATCCATTCGGAGGGCCATCTGCATTTACGGGGCGGGGACGCTGACTGGATTCTCCAGAA	3835
Qy	1272	ccatcacccgtttggccaaagaatttgaacccaagaagtatagataatagctctctcagtcct	1331
Db	3836	CCATACACCTCTTGGCAAAAGTTTATACCCGAGGGGTGATTAACGTCCTGCGAGTGGT	3895
Qy	1332	agaagaatacctaagctatataaggccatctatbtgagctgaactttgagacacatgata	1391
Db	3896	AGAGAGTACTACTACGTATCATTAGGCCATCTATATGGAGTCAACTTTTAGCTGCACACTGT	3951
Qy	1392	gcgttgctaacctgttgggcaagagacatgtacaaacaatgctcagaactgtctggagacagt	1451
Db	3952	-----GGGCGAGGAGCTGTGCCAACATGTGCGGCGGCGTACAGTACGTG	3996
Qy	1452	t-ggggtggagaccagggctttgcaatttcgttggcatctcttaagataaagggtcgtctat	1510
Db	3997	TGGGGTGGGGAGAGAGGCTGTGGAATTCATGTGATCCCCAGGGTCAAGAGGCTGCTTT	4056
Qy	1511	tagatttgggtgaatag 1528	
Db	4057	GATGGGGTCAAGATTTTG 4074	

RESULT	13
LOCUS	BC018297
DEFINITION	BC018297 1841 bp mRNA, linear ROD 06-DEC-2001 Mus musculus, RIKEN cDNA 2010013H2 gene, clone MGC:18826
IMAGE	4208500, mRNA, complete cds.
ACCESSION	BC018297

VERSION BC018297.1 GI:17390702
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1841)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalobedcm.tmc.edu
 Villalob, D.K., Luna, R.A., Hale, S.M., Huijck, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 24 Row: P Column: 13
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Similarity but not identity to protein.
 Location/Qualifiers
 1..1841
 /organism="Mus musculus"
 /db_xref="locusID:72077"
 /db_xref="taxon:10090"
 /clone="MGC:18826 IMAGE:4208500"
 /tissue_type="Colon, normal, 5 month old male mouse."
 /clone_1lb="NCL CGAP_C024"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 192..1388
 /codon_start=1
 /product="Riken cDNA 2010013H22 gene"
 /protein_id="AAH18297.1"
 /db_xref="GI:17390703"
 /translation="MDLDSEEFQSQYCRDLIYKTLKLPKASINCSGVINEQKAVTQC
 ALNNLEIKRKOQLFTEDADYLRMTADESHEPTKRRFLTIOSEASPIIAYSVVHIE
 KIENFERLRVAVTPPOVNYCVHMOCKSEPPKQVRAIVSCFPVFIASKVSVVYVH
 WSRQADLNCMEDLDLOSPPVWKVYLLNTGCTGPIKTVNAEYKALKLKGQMSSEVAP
 PPKHSKMKRYHEVTDLLHMTSKRRTPPNNLTMTGNAYVVASRDFEHVFSMSKAR
 QLTWVMDYVSPDEHIAVTLQRAISMGSGDPIHRKPOLSDMRALIAPLTKWYDHGCDIE
 NGATVYSCSGIHQRAVYVYSGSDLHWIILQNHLLANKFPDVPVDNVNLOCLEEYLRHKRA
 IYGEEL"
 BASE COUNT 496 a 448 c 469 g 428 t
 ORIGIN
 Query Match 49.0%, Score 923.8; DB 10; Length 1841;
 Best Local Similarity 73.1%; Pred. No. 1.le-283;
 Matches 1307; Conservative 0; Mismatches 447; Indels 35; Gaps 8;
 0Y 46 tccattctgtgagagatgttcaatggaagagacctctgcagctgcatctactgttgaggt 105
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 60 tccctgcctctggaagatgactcttcggagagagcctctgctgacactatcgccttgagacc 119
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 0Y 106 ctggagctctatatactctgcagcaccttgctctggaacctcttcaggttgaagtgt 165
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 120 ctgggtctctacatgctactgagcattccttgccttggaactgtctcctgagactgaagtgt 179

Oy 166 gactctgaccacttgggtctctgagtcacagggaattcaaaagccagttactgttagaataac 225
 Db 180 GACTTCGATGCGCATGATCTGGACTCTGAGAAATTTCAAAGCCACTACTGCGAGGATCTC 239
 Oy 226 ttgttaattccctgaacctccagaagaggtctataactgttccagggtcccccga 285
 Db 240 CTGTACAAAGACCCCTTAAGTTGCGACCAAGAAAGTTCCATCACTGCTCAGGGGTCTTTCGA 299
 Oy 286 ggggaccaaagagcagctgtcttcaagctattctgaaataactgttgaagagagagca 345
 Db 300 GGGGAGCAAGAAAGCGGTGACCCAGGCTCTGCTGATTAACCTTGAATTAAGAAAGAACGAG 359
 Oy 346 gagccttcaagagacacccactactctccctcaccagaagactgttgaacttcaaggct 405
 Db 360 CAGCTCTTACACAGGCGGACTACTTCTAGATGACAGCAGCACTGTGACACTTTCAGAAC 419
 Oy 406 gaaagaagttcatatcaagttccacttggcacaagaagaggttccctattgcatac 465
 Db 420 AAGAGGAAGTTTATACAGTCCACTGAGCAAGAGAGGCGCAGCTTCCCAATTCGCTAC 479
 Oy 466 tctatgtgattcatgagaagaattgaaaacttgaagaagctacttggagctgtgtatgc 525
 Db 480 TCCATGTTGTCATGATGAGAAATTGAGAACTTCGAAAGGTTCTGCGAGCTGTGACAC 539
 Oy 526 cctcagaacatactactgtgtccatgttggatgagaagtcctccagaacttcaagaagcgc 585
 Db 540 CCTCAGAAATGATGATGTCACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599
 Oy 586 gtcaagaacataattcttctgtctcccaaatgtcttcatatagcagtaagctgttgcggtg 645
 Db 600 GTCAGAGGCACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
 Oy 646 gtttctgctctgttctcagaaggttgaaggttgaaggttgaaggttgaaggttgaaggttga 705
 Db 660 GTCTATGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719
 Oy 706 agctcagtcgctgtggaataacttctcgaatacatgttggagcagacttctcctaagaagc 765
 Db 720 AGCCCGTGGCATGGAATTAACCTCTGAAACCTGTGGGACAGACTTTCCTCAAAACC 779
 Oy 766 aatgcagaagatgttccaggtctcagaatgttgaatgttgaatgttgaatgttgaatgttga 825
 Db 780 AATGCTGAGATGCTCAAGGCCCTCAAGCTATTGAAGGCGACACAGATGAGAGTCAGAG 839
 Oy 826 gtactctcgaagcacaagaacccgctcgaagaatactacttgaaggttgaaggttgaaggttga 885
 Db 840 GTACCACTCCCATTAATAAAATCCCGCTGGAATATCATATGAG--GTGACACACACA 896
 Oy 886 ttacaccttaacaaagaagaagagatccctcccttataacttataacttataacttataacttata 945
 Db 897 TTGCACATGACCGACGAG 956
 Oy 946 aatgcgtatcatgttgccttccgagattgtcacaacatgttgaagaagacccataatcc 1005
 Db 957 AATGCTCATATGCTGCTCTCTGACAGCTTTCATTTGAACAGCTTTCATTTGAACAGAGAGCC 1016
 Oy 1006 caacaacgtatgaaatgttgaagaagacacttataagcagaagacacttataagcagaagc 1065
 Db 1017 CGGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1076
 Oy 1066 ctctcagctgtcagagtgatgtcgtgtctgttctcacaacaccccaagtaagacatactca 1125
 Db 1077 CTCACAGCGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1136
 Oy 1126 gacatgaacttataatgtcagagctgtcgaagtgtcagaggttcatgtgagagagacatactaa 1185
 Db 1137 GACATGAGAGCGATGTCGAGACTTAACCAAGTGTGACGACATGAGAGAGACATTTAGAAC 1196
 Oy 1186 ggtgtctctatgtccctcgtctgtgataccacagggggtcatctgttgttattgtggct 1245
 Db 1197 GGGGACCTTTACACGCTTTGCTCAAGAAATCCACAGCGGGGCTGTGTGTATTTATGGGTCA 1256
 Oy 1246 ggggacttgaattgtgtcttcaaaacataccactgttggccaacaagtlttgaccaagaag 1305

Db 1257 GGGGACCTGCACTGATGATCTTCAAGACATCACCTCTTGGCCAAAGATTTGACCAACAG 1316
 Oy 1306 ttagatgaatgtctctcagtgcttgaagaatactcactgttataaggtcactataggg 1365
 Db 1317 GTGATGATTAATGTTCTTCAAGCTTTAGAGAAATTTTACGTACACAAAGCCATCTATGGG 1376
 Oy 1366 actgactttagacacacataagagcgttctactccttggggcaagacatgatacaaa 1425
 Db 1377 ACTGAACATATGAG--ACACTGAACATATTTGTTACTGCTGTGGCAAGAAATG----- 1428
 Oy 1426 catgtcagaagactgtcgtgagcagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1485
 Db 1429 -----CCAGGACCTTACTGGGCGAGTGTGGGAGATACCTGGGGCTCTAGACTCCATGTG 1483
 Oy 1486 tcccttgaataagagggctgtatgaattgttggtagaagacttcttgccttgcacaa 1545
 Db 1484 TCCCTTGAATGAGGGGCTGTATTTGACACGGGAGTAGTACCTGCTGCTCTTGTGA 1543
 Oy 1546 ttgtcgtcgt 1605
 Db 1544 --GCTGTGCTGT 1594
 Oy 1606 acttctcactaagatgagaatgagaacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1661
 Db 1595 GCTTCTCGCAGAGCTAGACATGAATATCTGTACTTGGGATATTAAGGAGGGGAGACAT 1654
 Oy 1662 atgtgtgtgagacacttgaattcagttgaatgcctgtgtgtgtgtgtgtgtgtgtgtgtgt 1721
 Db 1655 GTGACCAAGGACCCCTGT 1709
 Oy 1722 ctgcgcctcctaataatcagaagttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1781
 Db 1710 ACTTGTGTTCAATACCTCTGCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1769
 Oy 1782 ccttcccttctgtactgttacttataaataaataagcttcttgcattcaaa 1830
 Db 1770 CCCCCTCTTTTATTTATTTTAA--TTTAACTATAAGAGCTTCTTAACAAA 1817

RESULT 14
 AC096432/C
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-36K24, *** SEQUENCING IN PROGRESS
 *** 64 unordered pieces.
 ACCESSION
 AC096432.4 GI:17946738
 VERSION
 AC096432.4
 KEYWORDS
 HTG: HTGS_PHASE1.
 SOURCE
 Norway rat.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE
 1 (bases 1 to 176483)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oshan,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarta,D.,
 Benton,J., Bivame,K., Blankenburg,K., Bonini,D., Bouck,J.,
 Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,J., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Geo,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Hollway,C.,
 Hollins,B., Homs,I.F., Howard,S., Huber,J., Huliyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Deal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W., Louised, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkwo, S., Ogun, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Quidley, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 176483)
Moriarty, K.C.

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062249.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GPRP
Center clone name: CH230-36K24
----- Summary Statistics
findhaplaid
Assembly program: Phrap; version 0.990329First call to
findhaplaid
Consensus quality: 145213 bases at least Q40
Consensus quality: 154367 bases at least Q30
Consensus quality: 161818 bases at least Q20
Estimated insert size: 148877; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-IP estimation
Quality coverage: 2.1x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
10789: contig of 10789 bp in length
10790 10889: gap of unknown length
10890 17590: contig of 6701 bp in length
17591 17690: gap of unknown length
17691 23617: contig of 5927 bp in length
23618 23717: gap of unknown length
23718 27572: contig of 3855 bp in length
27573 27672: gap of unknown length
27673 31919: contig of 4247 bp in length
31920 32019: gap of unknown length
32020 37411: contig of 5392 bp in length
37412 37511: gap of unknown length
37512 41536: contig of 4025 bp in length
41537 46655: gap of unknown length
46656 46755: contig of 5019 bp in length
46756 46755: gap of unknown length

51783: contig of 5028 bp in length
51784 51883: gap of unknown length
51884 55562: contig of 3679 bp in length
55563 55662: gap of unknown length
55663 59942: contig of 4280 bp in length
59943 60042: gap of unknown length
60043 63342: contig of 3300 bp in length
63343 63442: gap of unknown length
63443 69023: contig of 5581 bp in length
69024 73096: contig of 3973 bp in length
73097 73196: gap of unknown length
73197 77220: contig of 4024 bp in length
77221 77320: gap of unknown length
77321 80222: contig of 2902 bp in length
80223 80332: gap of unknown length
80333 84643: contig of 4321 bp in length
84644 84743: gap of unknown length
84744 87894: contig of 3151 bp in length
87895 87994: gap of unknown length
87995 89819: contig of 1825 bp in length
89820 89919: gap of unknown length
89920 93241: contig of 3322 bp in length
93242 93341: gap of unknown length
93342 95561: contig of 2220 bp in length
95562 95661: gap of unknown length
95662 98293: contig of 2632 bp in length
98294 98393: gap of unknown length
98394 99649: contig of 1256 bp in length
99650 99749: gap of unknown length
99750 102029: contig of 2280 bp in length
102030 102129: gap of unknown length
102130 105439: contig of 3310 bp in length
105440 105539: gap of unknown length
105540 107561: contig of 2022 bp in length
107562 107661: gap of unknown length
107662 110897: contig of 3236 bp in length
110898 110997: gap of unknown length
110999 114069: contig of 3072 bp in length
114100 114169: gap of unknown length
114170 116366: contig of 2197 bp in length
116367 116466: gap of unknown length
116467 118335: contig of 1865 bp in length
118336 118435: gap of unknown length
118436 120586: contig of 2151 bp in length
120587 120686: gap of unknown length
120687 121822: contig of 1136 bp in length
121823 121922: gap of unknown length
121923 124018: contig of 2056 bp in length
124019 124118: gap of unknown length
124119 125819: contig of 1701 bp in length
125820 125919: gap of unknown length
125920 127438: contig of 1519 bp in length
127439 127538: gap of unknown length
127539 129346: contig of 1808 bp in length
129347 129446: gap of unknown length
129447 132165: contig of 2719 bp in length
132166 132265: gap of unknown length
132266 134061: contig of 1796 bp in length
134062 134161: gap of unknown length
134162 136403: contig of 2242 bp in length
136404 136503: gap of unknown length
136504 138291: contig of 1788 bp in length
138292 138391: gap of unknown length
138392 140162: contig of 1771 bp in length
140163 140262: gap of unknown length
140263 142202: contig of 1940 bp in length
142203 142302: gap of unknown length
142303 144059: contig of 1757 bp in length
144060 144159: gap of unknown length
144160 145878: contig of 1719 bp in length
145879 145978: gap of unknown length
145979 147443: contig of 1465 bp in length

Noncoding region o
Human core 26Grt DN
cDNA sequence of h
Human heart core 2
Rat DH1 cDNA. Rat
Diabetic rat heart
Coding sequence fo
Human colon cancer
DNA encoding novel
Human DNA encoding
I-branchning enzyme
Beta-1,6-N-acetylgl
Full length cDNA s
Human beta-1,6-N-a
C2Grt catalytic do
DNA encoding novel
Human polynucleoti
cDNA encoding a re
Murine beta-1,6-N-
Murine beta-1,6-N-
Core 2 beta-1,6-N-
Noncoding region o
Novel human polyu
Human EST-derived
C2Grt C-terminal r
C2Grt C-terminal r
Human foetal liver
Human foetal liver
Human foetal liver
Human brain expres
Human bone marrow
Probe #11365 for g
Probe #15407 used f
Probe #7118 used t
Human foetal liver
Probe #2154 for ge
Human brain expres

Dh 1996 atgctgctgttctctcaacccctaaacctagtagtctcctcaactaacttctcactaagt 2055
Qy 1621 gagaatggaactgctgtgataaggagatgaaggagatatgtgtagagacttgat 1680
Dh 2056 gagaatggaactgctgtgataaggagatgaaggagatatgtgtagagacttgat 2115
Qy 1661 ttcaagtgaatgctgtgtgtagctttccattctgtgtgagctgcgttccctaataatc 1740
Dh 2116 ttcaagtgaatgctgtgtgtagctttccattctgtgtgagctgcgttccctaataatc 2175
Qy 1741 caggttgtagctgtgtgagaggaacttgatgtaagaagaacctccctctgtactgtt 1800
Dh 2176 caggttgtagctgtgtgagaggaacttgatgtaagaagaacctccctctgtactgtt 2235
Qy 1801 aacttaaaataaagctcctgactcaagaagtattacccttactcttctgacagatgcc 1860
Dh 2236 aacttaaaataaagctcctgactcaagaagtattacccttactcttctgacagatgcc 2295
Qy 1861 agaataataataalctaaacaga 1884
Dh 2296 agaataataataalctaaacaga 2319

RESULT 2

AAA96570 standard; DNA; 2108 BP.

AAA96570;

08-FEB-2001 (first entry)

A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.

Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
cancer; cardiovascular disorder; inflammatory disorder; asthma;
rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
septic shock; adult respiratory distress syndrome; ARDS; cancer;
platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
diverticulitis; ulcerative colitis; ss.

Homo sapiens.

CA2296936-A1.

03-AUG-2000.

03-FEB-2000; 2000CA-2296936.

03-FEB-1999; 99US-0118674.

(GLYC-) GLYCODESIGN INC.

Korczak B, Lew A;

WP1; 2000-594746/57.

New nucleic acid molecules of core 2

beta-1,6-N-acetylglucosaminyltransferase useful for providing new

compositions for treatment of disorders mediated by the enzyme

including cancer, cardiovascular and inflammatory disorders.

Claim 4; Page 51-52; 66pp; English.

The present sequence encodes a partial human core 2

beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.

The polypeptide can be used to treat diseases and disorders, such as

cancer, cardiovascular disorders and inflammatory disorders including

asthma, rheumatoid arthritis, inflammatory bowel disease,

arteriosclerosis, septic shock, adult respiratory distress syndrome

CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.

Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;

Query Match 97.3%; Score 1833; DB 21; Length 2108;

Best local similarity 100.0%; Pred. No. 0;

Matches 1833; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gattgtgctccctccctacccttccctgtgtctgtgtccactgtgtccattctgtgacg 60
Dh 208 gattgtgctccctccctacccttccctgtgtctgtgtccactgtgtccattctgtgacg 267
Qy 61 atgtgtcaatggaagagactgtccagctgcatctactgtgtggtctgtgtatag 120
Dh 268 atgtgtcaatggaagagactgtccagctgcatctactgtgtggtctgtgtatag 327
Qy 121 ctgtctgccaactgtgtctctgaaacttcttcaagttgaagtgtgactctgacacttg 180
Dh 328 ctgtctgccaactgtgtctctgaaacttcttcaagttgaagtgtgactctgacacttg 387
Qy 181 ggtctggaagtcgaaggaatctcaagccagctgtaggaatatctgtataattctctg 240
Dh 388 ggtctggaagtcgaaggaatctcaagccagctgtaggaatatctgtataattctctg 447
Qy 241 aaacttcagaagaaggtctcatcaactgttcaagggtgcaccggaggaaccaagaagca 300
Dh 448 aaacttcagaagaaggtctcatcaactgttcaagggtgcaccggaggaaccaagaagca 507
Qy 301 gtgtctcaggtatctgtaataacctggaaggtlcaagaagaagcgagccttcaagac 360
Dh 508 gtgtctcaggtatctgtaataacctggaaggtlcaagaagaagcgagccttcaagac 567
Qy 361 acccactaccctccctccctaccagaagactgtgagcacttcaagggtcgaaagaagtctca 420
Dh 568 acccactaccctccctccctaccagaagactgtgagcacttcaagggtcgaaagaagtctca 627
Qy 421 cagttccacatggaagaagaaggtgaggttccctattgtacatctatgtgtatcat 480
Dh 628 cagttccacatggaagaagaaggtgaggttccctattgtacatctatgtgtatcat 687
Qy 481 gagaagattgaaaccttgaaaggctactgcgagctgtgtatgccccacagaacataac 540
Dh 688 gagaagattgaaaccttgaaaggctactgcgagctgtgtatgccccacagaacataac 747
Qy 541 tgtgtccatgtggaagaagaagtcaccagaagaacttcaagaaggcggttaagaacatttt 600
Dh 748 tgtgtccatgtggaagaagaagtcaccagaagaacttcaagaaggcggttaagaacatttt 807
Qy 601 tctgtctcccaaatgtcttcatatagccagtaagctgtgtcggtgtgttatgtctctcg 660
Dh 808 tctgtctcccaaatgtcttcatatagccagtaagctgtgtcggtgtgttatgtctctcg 867
Qy 661 tccaaggtgcaagctgacatcaactgcatggaagaactgtgtccagagctcagtcggtg 720
Dh 868 tccaaggtgcaagctgacatcaactgcatggaagaactgtgtccagagctcagtcggtg 927
Qy 721 aaatactctcgaataactgtggaagcgagacttctctataaagagaatgcaagatggtc 780
Dh 928 aaatactctcgaataactgtggaagcgagacttctctataaagagaatgcaagatggtc 987
Qy 781 caggtctccaagatgttaatgggaggaatagcatgaggtcaagagttaccctccaagac 840
Dh 988 caggtctccaagatgttaatgggaggaatagcatgaggtcaagagttaccctccaagac 1047
Qy 841 aaagaacccgctggaataatcactttaggttagtgaagaacacattacactaaccaac 900

Db 1048 aaagaaccgctggaatactcaattgagtgtagagagacacattacacctaaccac 1107
 QY 901 aagaagaaggatcctcccttaataattactatgtttacagggaatgctacattgtg 960
 Db 1108 aagaagaaggatcctcccttaataattactatgtttacagggaatgctacattgtg 1167
 QY 961 gcttcccgagatttcttccacaatgttttgaagaacccctaataatccacaactgattgaa 1020
 Db 1168 gcttcccgagatttcttccacaatgttttgaagaacccctaataatccacaactgattgaa 1227
 QY 1021 tgggttaaaagacattatagcccaagatgaacacctctggggccaccccttcagctgcacgg 1080
 Db 1228 tgggttaaaagacattatagcccaagatgaacacctctggggccaccccttcagctgcacgg 1287
 QY 1081 tggatgcttggctctgttcccaacaccccaagtaacacatcctcaagatgacttctatt 1140
 Db 1288 tggatgcttggctctgttcccaacaccccaagtaacacatcctcaagatgacttctatt 1347
 QY 1141 gccaggtgttcaagtggcaggggtcatgagggagacatgataagggtgcttcttattgtc 1200
 Db 1348 gccaggtgttcaagtggcaggggtcatgagggagacatgataagggtgcttcttattgtc 1407
 QY 1201 cccttgccttgaatcccaacgagcgtatctgcgtttatgtgggctgggagacttgaattgg 1260
 Db 1408 cccttgccttgaatcccaacgagcgtatctgcgtttatgtgggctgggagacttgaattgg 1467
 QY 1261 atgcttcaaaacatcacctgttggccaacaagtttgacccaagtagatgaatgact 1320
 Db 1468 atgcttcaaaacatcacctgttggccaacaagtttgacccaagtagatgaatgact 1527
 QY 1321 ctccaggtctagaagaatcctcactatgaagggcactatgagtgagtgacttgacttgagac 1380
 Db 1528 ctccaggtctagaagaatcctcactatgaagggcactatgagtgagtgacttgagac 1587
 QY 1381 acaatagagaagcgttctcaccctgtgggcaagaagcactgtaacaacatgtctcgaacttg 1440
 Db 1588 acaatagagaagcgttctcaccctgtgggcaagaagcactgtaacaacatgtctcgaacttg 1647
 QY 1441 ctgggagacgttgggtgggagacaggcgtttgcaattcgttgcatcctttagataaga 1500
 Db 1648 ctgggagacgttgggtgggagacaggcgtttgcaattcgttgcatcctttagataaga 1707
 QY 1501 gggctgtatagaatgttggttagatgactcttgccttgcgaatattgctgcctgggtga 1560
 Db 1708 gggctgtatagaatgttggttagatgactcttgccttgcgaatattgctgcctgggtga 1767
 QY 1561 atgctgttcttctcaccaccccaacccatgtagttcctcactaacttctcactaagt 1620
 Db 1768 atgctgttcttctcaccaccccaacccatgtagttcctcactaacttctcactaagt 1827
 QY 1621 gagaatagaagactgtgtgaataggagagatgaaggagagatatgtgtgtagagcaattgat 1680
 Db 1828 gagaatagaagactgtgtgaataggagagatgaaggagagatatgtgtgtagagcaattgat 1887
 QY 1681 ttaagtgaatgctgtgtgtgaatttccattctgtgagcgctgttccataatc 1740
 Db 1888 ttaagtgaatgctgtgtgtgaatttccattctgtgagcgctgttccataatc 1947
 QY 1741 caggttggagcgttggagagagacttgaatggaaagagaaaccttccctcttactggt 1800
 Db 1948 caggttggagcgttggagagagacttgaatggaaagagaaaccttccctcttactggt 2007
 QY 1801 aacttaaaaataatagctccttgaatcaagta 1833
 Db 2008 aacttaaaaataatagctccttgaatcaagta 2040

RESULT 3
 AAH34463
 ID AAH34463 standard; cDNA; 2236 BP.
 XX
 AC AAH34463;
 XX

DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
 XX
 KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW Colorectal carcinoma; chromosome 15; 5S.
 XX
 OS Homo sapiens.
 PN W0200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000MO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI: 2001-235357/24.
 DR P-PSDB; AAG75058.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PS Claim 1; Page 3190-3191; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have collective activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other;

Query Match 97.3%; Score 1833; DB 22; Length 2236;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 1833; Conservative 0;

QY 1 gatttgctcctccacaccttccctgtgctggtctccacactgtcccatctgtgagc 60
 Db 374 gatttgctcctccacaccttccctgtgctggtctccacactgtcccatctgtgagc 433
 QY 61 atggttcaatggaagactctgcagctgacattctgtgggtctgggctgtatattg 120
 Db 434 atggttcaatggaagactctgcagctgacattctgtgggtctgggctgtatattg 493
 QY 121 ctgctggcactgtgctctggaacttctttaggttgaagtgtaacctgacccactg 180
 Db 494 ctgctggcactgtgctctggaacttctttaggttgaagtgtaacctgacccactg 553
 QY 181 ggtctggaatcagggaatctcaagcagctactgttagagatatctgtataattctcgt 240
 Db 554 ggtctggaatcagggaatctcaagcagctactgttagagatatctgtataattctcgt 613

QY 241 aaacttcagcaagaagatctatactatgttcagggtccaccggagggagccaagaagca 300
|||||
Db 614 aaacttcagcaagaagatctatactatgttcagggtccaccggagggagccaagaagca 673
QY 301 gtgtctcagggtatctctgaataacctggaaggtcgaagaagcgagagctttccagac 360
|||||
Db 674 gtgtctcagggtatctctgaataacctggaaggtcgaagaagcgagagctttccagac 733
QY 361 acccctcctctccctcccaacgagactgtgagcacttaagagctcgaaagagttcata 420
|||||
Db 734 acccctcctctccctcccaacgagactgtgagcacttaagagctcgaaagagttcata 793
QY 421 cagttcccaactgagcaaaagaggtgaggtccctatgcatactctatgttgattcat 480
|||||
Db 794 cagttcccaactgagcaaaagaggtgaggtccctatgcatactctatgttgattcat 853
QY 481 gagaagatgaaaccttgaagaagctcgaagctgtgtatgtccctcagaacataaac 540
|||||
Db 854 gagaagatgaaaccttgaagaagctcgaagctgtgtatgtccctcagaacataaac 913
QY 541 tgtgtccatgtgagtagagaggtccccaagaacttccaagaagcggtccaagaacttat 600
|||||
Db 914 tgtgtccatgtgagtagagaggtccccaagaacttccaagaagcggtccaagaacttat 973
QY 601 tctgtctcccaaatgtctcctcaagcaagctgtgtcgggtgtttatgtcctctcg 660
|||||
Db 974 tctgtctcccaaatgtctcctcaagcaagctgtgtcgggtgtttatgtcctctcg 1033
QY 661 tccaggtgtgcaagctggaactcgaactgtgcatggaagactgtctccagaagctcagtgccgtg 720
|||||
Db 1034 tccaggtgtgcaagctggaactcgaactgtgcatggaagactgtctccagaagctcagtgccgtg 1093
QY 721 aaatctctctgaataacatggtggaagctcttccataaagaagatgcaagatgtgc 780
|||||
Db 1094 aaatctctctgaataacatggtggaagctcttccataaagaagatgcaagatgtgc 1153
QY 781 caggtctccaagaatgttgaatggaagaaatagcatgagtcagaggtacctcctaagcac 840
|||||
Db 1154 caggtctccaagaatgttgaatggaagaaatagcatgagtcagaggtacctcctaagcac 1213
QY 841 aaagaacccgctgtgaaataatacaacttgaagtgagagacacattatacctaacaac 900
|||||
Db 1214 aaagaacccgctgtgaaataatacaacttgaagtgagagacacattatacctaacaac 1273
QY 901 aagaagaagaatcccccctataatatactatgtttacaggggaatcgtaacattgtg 960
|||||
Db 1274 aagaagaagaatcccccctataatatactatgtttacaggggaatcgtaacattgtg 1333
QY 961 gctcccgagatttcgtccacaacagtlttgaagaaccttaaaatcccaacaactgtatga 1020
|||||
Db 1334 gctcccgagatttcgtccacaacagtlttgaagaaccttaaaatcccaacaactgtatga 1393
QY 1021 tgggttaaaagacattatagccagatgaacacctgtggccaacctttagagtgtagcg 1080
|||||
Db 1394 tgggttaaaagacattatagccagatgaacacctgtggccaacctttagagtgtagcg 1453
QY 1081 tggatgacctgtctgttcccaacaccccaagtaagacatctcaagatgatttat 1140
|||||
Db 1454 tggatgacctgtctgttcccaacaccccaagtaagacatctcaagatgatttat 1513
QY 1141 gccaggtgtgtcaagtgtgcaaggtcagtgtatgttgagggctgtggagacttgatgtg 1200
|||||
Db 1514 gccaggtgtgtcaagtgtgcaaggtcagtgtatgttgagggctgtggagacttgatgtg 1573
QY 1201 cctgtcttggaatccaccagcggtatctcggtttaggggctgtggagacttgatgtg 1260
|||||
Db 1574 cctgtcttggaatccaccagcggtatctcggtttaggggctgtggagacttgatgtg 1633
QY 1261 atggtctcaaaacacacatctgtgtgccaacaagtttgaccaagaagtatgataatgt 1320
|||||
Db 1634 atggtctcaaaacacacatctgtgtgccaacaagtttgaccaagaagtatgataatgt 1693
QY 1321 ctctcagtgcttagaagataactacgttataagggcatctatgtggaactttagagac 1380

|||||
Db 1694 ctctcagtgcttagaagataactacgttataagggcatctatgtggagactttagagac 1753
QY 1381 acactatgagagcggtgtctaactctgtgggcaagagcatgtacaaacatgtctagaactg 1440
|||||
Db 1754 acactatgagagcggtgtctaactctgtgggcaagagcatgtacaaacatgtctagaactg 1813
QY 1441 ctgggacaggtgtgggttggaagccaggccttggcaatcgttgacatccttttagatga 1500
|||||
Db 1814 ctgggacaggtgtgggttggaagccaggccttggcaatcgttgacatccttttagatga 1873
QY 1501 gggctgtataatgactgtgtgggaatgatatcttctgtccttgcaaaatgtcgtctgggtga 1560
|||||
Db 1874 gggctgtataatgactgtgtgggaatgatatcttctgtccttgcaaaatgtcgtctgggtga 1933
QY 1561 atgtctgtgtctctcaaccctcaaccctagtagttctctcaactaacttctcaactagt 1620
|||||
Db 1934 atgtctgtgtctctcaaccctcaaccctagtagttctctcaactaacttctcaactagt 1993
QY 1621 gagaatgagaactgtctgtatagaggagagtggaagggagatatgtgtgtagagcaattgat 1680
|||||
Db 1994 gagaatgagaactgtctgtatagaggagagtggaagggagatatgtgtgtagagcaattgat 2053
QY 1681 ttcagttgaaatgcctgtgtgtagactttccatctctgtggaagctgcggttccataaatc 1740
|||||
Db 2054 ttcagttgaaatgcctgtgtgtagactttccatctctgtggaagctgcggttccataaatc 2113
QY 1741 caggtttgtgtagcgtgtggaagaaacttgaatggaagaagaaaccttccctctgtatcgt 1800
|||||
Db 2114 caggtttgtgtagcgtgtggaagaaacttgaatggaagaagaaaccttccctctgtatcgt 2173
QY 1801 aacttaaaataaataagctcctgattccaagta 1833
|||||
Db 2174 aacttaaaataaataagctcctgattccaagta 2206

RESULT 4
AAC99109
ID AAC99109 standard; cDNA; 2229 BP.
XX
AC AAC99109;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
XX
OS Homo sapiens.
XX
PN WO20005320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-579444/54.
XX
DR P-PSDB: AAB54344.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
treating, or ameliorating a medical condition, particular pancreatic

PT cancer, or for use in assays for diagnosing a pathological condition -
XX Claim 1; Page 759-760; 1379pp; English.
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neurotropic, immunomodulatory, relaxant, contractile,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SO Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;

Query Match 97.3%; Score 1832.6; DB 21; Length 2229;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1832; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatgtgtctctcccaaccccttcctgctgcgcggtcctcacctgtctccattctgtgacg 60
DB 367 gatgtgtctctcccaaccccttcctgctgcgcggtcctcacctgtctccattctgtgacg 426
QY 61 atggttcaatggaagagactctgcacgtcgtacattgttgggtctcgtgctgatatg 120
DB 427 atggttcaatggaagagactctgcacgtcgtacattgttgggtctcgtgctgatatg 486
QY 121 ctgtctgacactgtgtgtctggaacttcttcaaggttgaagtgtgactctgacacttg 180
DB 487 ctgtctgacactgtgtgtctggaacttcttcaaggttgaagtgtgactctgacacttg 546
QY 181 ggtctggaagccagaagatctcaaaagccagtactgtagatatctgtataattctctg 240
DB 547 ggtctggaagccagaagatctcaaaagccagtactgtagatatctgtataattctctg 606
QY 241 aaacttcagcaaaagaggtctatcaactgttcaagggttcacccgaggggaccaaaggca 300
DB 607 aaacttcagcaaaagaggtctatcaactgttcaagggttcacccgaggggaccaaaggca 666
QY 301 gtgttcaaggtctatcgaataacctgtgagggtcaagaagaagcgagagcctttcacagac 360
DB 667 gtgttcaaggtctatcgaataacctgtgagggtcaagaagaagcgagagcctttcacagac 726
QY 361 accacatacctctccctcacagaagactgtgagcaacttcaagggttgaagaagattcata 420
DB 727 accacatacctctccctcacagaagactgtgagcaacttcaagggttgaagaagattcata 786
QY 421 cagttccactgagcaaaagaaggtgaggttccctattgcatactatctatgtgtatcat 480
DB 787 cagttccactgagcaaaagaaggtgaggttccctattgcatactatctatgtgtatcat 846
QY 481 gagaagaattgaaaacttgaagaagctacgtcgagctgtgtaagccctcagaataatatac 540
DB 847 gagaagaattgaaaacttgaagaagctacgtcgagctgtgtaagccctcagaataatatac 906
QY 541 tgttccatgtgatagtgaaggtccccaagaacttcaaaagaagcggtcgaagaactattt 600
DB 907 tgttccatgtgatagtgaaggtccccaagaacttcaaaagaagcggtcgaagaactattt 966

QY 601 tcttgtctcccaaatgtcttcaatagccagtaagctgttcgggtgtttatctcctctg 660
DB 967 tcttgtctcccaaatgtcttcaatagccagtaagctgttcgggtgtttatctcctctg 1026
QY 661 tcaaggttgaagaactgacccaactgactgtgaagaagactgttccaaggtcgaagtcggtg 720
DB 1027 tcaaggttgaagaactgacccaactgactgtgaagaagactgttccaaggtcgaagtcggtg 1086
QY 721 aaatactctctgatacatgttggagcagacttccctataagaagcagatgagagatgtgtc 780
DB 1087 aaatactctctgatacatgttggagcagacttccctataagaagcagatgagagatgtgtc 1146
QY 781 caggcttcaagaatgttgaatgggaggaatagacatgagatcagaaggtacccctaaagcac 840
DB 1147 caggcttcaagaatgttgaatgggaggaatagacatgagatcagaaggtacccctaaagcac 1206
QY 841 aagaagaacccgctggaataatcaactttagagtgatgagagacatcaaccctaaaccaac 900
DB 1207 aagaagaacccgctggaataatcaactttagagtgatgagagacatcaaccctaaaccaac 1266
QY 901 aagaagaagatccctcccttaataattacatgtttacaggaatggtacatgtgtg 960
DB 1267 aagaagaagatccctcccttaataattacatgtttacaggaatggtacatgtgtg 1326
QY 961 gttcccgagattcgttccacaatgtttagaagaacccctaaatcccaacaactgattgaa 1020
DB 1327 gttcccgagattcgttccacaatgtttagaagaacccctaaatcccaacaactgattgaa 1386
QY 1021 tgggttaaaagacattatagccagatgaacacccctcgggacaccccttaagcgttgacgg 1080
DB 1387 tgggttaaaagacattatagccagatgaacacccctcgggacaccccttaagcgttgacgg 1446
QY 1081 tggatgcctgtgctgtgttcccaacccccaagtlacagacatcccaagacatgttctat 1140
DB 1447 tggatgcctgtgctgtgttcccaacccccaagtlacagacatcccaagacatgttctat 1506
QY 1141 gccaggtctgttcaagtgagcagggtcaltagaggagaacatcgataagggtgctctatgct 1200
DB 1507 gccaggtctgttcaagtgagcagggtcaltagaggagaacatcgataagggtgctctatgct 1566
QY 1201 cctctgctctggaatccacacgagcggtatcgtgttatgttgaggtcgtggagcttgaatttg 1260
DB 1567 cctctgctctggaatccacacgagcggtatcgtgttatgttgaggtcgtggagcttgaatttg 1626
QY 1261 atgcttcaaaaacatacccggttggccaacaagtttgaccaacaaggtagataatgtct 1320
DB 1627 atgcttcaaaaacatacccggttggccaacaagtttgaccaacaaggtagataatgtct 1686
QY 1321 ctccaagtgcttaagaagaatactacgttataagggcaactctatgttgagacttgagac 1380
DB 1687 ctccaagtgcttaagaagaatactacgttataagggcaactctatgttgagacttgagac 1746
QY 1381 acactatgagaagcgtgtgtctacccctgttgggcaagaagcatgtacaacaatgtcagaacttg 1440
DB 1747 acactatgagaagcgtgtgtctacccctgttgggcaagaagcatgtacaacaatgtcagaacttg 1806
QY 1441 ctgggaacagtgtgtgtggagacagaggtcttgcaattctgttgcaccccttaagataaga 1500
DB 1807 ctgggaacagtgtgtgtggagacagaggtcttgcaattctgttgcaccccttaagataaga 1866
QY 1501 gggctgtatataatgtgtgtgtgaatagatccttgccttgcaaatgtcgtcgtgtgtga 1560
DB 1867 gggctgtatataatgtgtgtgtgaatagatccttgccttgcaaatgtcgtcgtgtgtga 1926
QY 1561 atgctgtgttcttccacccctaaacccctagtaggttccctccataacttctcaactaagt 1620
DB 1927 atgctgtgttcttccacccctaaacccctagtaggttccctccataacttctcaactaagt 1986
QY 1621 gagaatgaagacgtcgtgtgaataggaagatgaaggaaggaatacgtgtgtagagacacttgat 1680
DB 1987 gagaatgaagacgtcgtgtgtgaataggaagatgaaggaaggaatacgtgtgtagagacacttgat 2046

QY	1081	tgagatgcctggtctctgttcccaaccaccccaagtagacatctcagacatgacttctatt	1140
Db	1374	tgagatgcctggtctctgttcccaaccaccccaagtagacatctcagacatgacttctatt	1433
QY	1141	gcacgagctggtcgaagtggcagggctcatggaggagacatcgataaaggtgtcccttatgct	1200
Db	1434	gcacgagctggtcgaagtggcagggctcatggaggagacatcgataaaggtgtcccttatgct	1493
QY	1201	cccgctcttggaatccacccacggcggtcattctggttttagtgggctgggagcttgtaatgg	1260
Db	1494	cccgctcttggaatccacccacggcggtcattctggttttagtgggctgggagcttgtaatgg	1553
QY	1261	atgcttcaaaaccatcacctgtgtggccaacaagttgaccccaaaagtatgatatgataatgct	1320
Db	1554	atgcttcaaaaccatcacctgtgtggccaacaagttgaccccaaaagtatgatatgataatgct	1613
QY	1321	cttaagtgcttagaagaagatcctcagtttaatgaagccattataggagactggaactttggac	1380
Db	1614	cttaagtgcttagaagaagatcctcagtttaatgaagccattataggagactggaactttggac	1673
QY	1381	acaactatgagagcgttctcaccgtctggggccaagaagcatgtacaacaatgctccaagaactg	1440
Db	1674	acaactatgagagcgttctcaccgtctggggccaagaagcatgtacaacaatgctccaagaactg	1733
QY	1441	ctggagacagtgtyggtygggagaccagggcttggcaattctgtggtacaccttaagataaga	1500
Db	1734	ctggagacagtgtyggtygggagaccagggcttggcaattctgtggtacaccttaagataaga	1793
QY	1501	gggcgcgcattatagattctgtggtagaagatcctttgccttcgaattctgctcgtgggtga	1560
Db	1794	gggcgcgcattatagattctgtggtagaagatcctttgccttcgaattctgctcgtgggtga	1853
QY	1561	atgctgctgttctctccaccctcaaccctagtagttctctccactaacttctccactaagt	1620
Db	1854	atgctgctgttctctccaccctcaaccctagtagttctctccactaacttctccactaagt	1913
QY	1621	gagagatgagaaactctgtgtgtatggaggagtggaaggagggatattgtgtgtagagcaactg	1680
Db	1914	gagagatgagaaactctgtgtgtatggaggagtggaaggagggatattgtgtgtagagcaactg	1973
QY	1681	ttcagttggaatgcctctgtgttagctttccattctcgtctggagctgccttccctaatactc	1740
Db	1974	ttcagttggaatgcctctgtgttagctttccattctcgtctggagctgccttccctaatactc	2033
QY	1741	cagggttggtagcgttggagagagaacttctgatgnaaagagaaccttccctctgtactgtt	1800
Db	2034	cagggttggtagcgttggagagagaacttctgatgnaaagagaac--ttcctctgtactgtt	2091
QY	1801	aacttaaaaaataaata 1816	
Db	2092	aacttaaaaaataaata 2107	
RESULT 6			
AAA96569			
ID	AAA96569 standard; DNA: 1317 BP.		
XX	AAA96569;		
AC			
XX			
DT	08-FEB-2001 (first entry)		
DE	DNA encoding a core 2 beta-1,6-N-acetylglucosaminyltransferase.		
XX			
KM	Core 2 beta-1,6-N-acetylglucosaminyltransferase: core2b GlcNAc-T;		
KM	cancer; cardiovascular disorder; inflammatory disorder; asthma;		
KM	rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;		
KM	septic shock; adult respiratory distress syndrome; ARDS; cancer;		
KM	platelet-mediated pathology; atherosclerosis; gastrointestinal disorder		
KM	clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;		
XX	diverticulitis; ulcerative colitis; ss.		
OS	Homo sapiens.		

xx	Key	Location/Qualifiers
PH	1..1317	
FT	CDS	
FT	/tag= a	
FT	/product=	"core 2
FT		beta-1,6-N-acetylglucosaminyltransferase"
xx		
PN	CA2296936-AL.	
xx		
XX	03-AUG-2000.	
xx		
PF	03-FEB-2000; 2000CA-2296936.	
xx		
PR	03-FEB-1999; 99US-0118674.	
xx		
PA	(GLYC-) GLYCODESIGN INC.	
xx		
PI	Korczak B, Lew A;	
xx		
DR	WPI. 2000-594746/57.	
xx	P-PSDB; AAB18995.	
xx		
PT	New nucleic acid molecules of core 2	
PT	beta-1,6-N-acetylglucosaminyltransferase useful for providing new	
PT	compositions for treatment of disorders mediated by the enzyme	
PT	including cancer, cardiovascular and inflammatory disorders.	
xx		
PS	Claim 4; Page 50; 66pp; English.	
xx		
CC	The present sequence encodes a human core 2	
CC	beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.	
CC	The polypeptide can be used to treat diseases and disorders such as	
CC	cancer, cardiovascular disorders and inflammatory disorders including	
CC	asthma, rheumatoid arthritis, inflammatory bowel disease,	
CC	arteriosclerosis, septic shock, adult respiratory distress syndrome	
CC	(ARDS) and cancer. Various platelet-mediated pathologies such as	
CC	atherosclerosis and clotting can also be treated. The polypeptides of	
CC	the invention are predominantly expressed in gastrointestinal tissue	
CC	(stomach, colon, intestine, testis) and are elevated in cancer.	
CC	Gastrointestinal disorders that may be prevented or treated include	
CC	ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis	
CC	and ulcerative colitis. The antibodies may be used in	
CC	immuno-histochemical analysis, to detect the novel polypeptide and to	
CC	localize it to particular cells and tissues and to specific subcellular	
CC	locations and to quantitate the level of expression.	
xx		
SQ	Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;	
Query Match	69.9%; Score 1317; DB 21; Length 1317;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1317; Conservative	0; Mismatches 0; Indels 0; Gaps	0;
QY	61 atggttcaatgtaagaagactctgccagctgcatacttggcgtcgtgcgtcatatg 120	
DB	1 atggttcaatgtaagaagactctgccagctgcatacttggcgtcgtgcgtcatatg 60	
QY	121 ctgctgccaactgtggtctgtaacttcttcagggtgaagtgtagtctgcacacttg 180	
DB	61 ctgctgccaactgtggtctgtaacttcttcagggtgaagtgtagtctgcacacttg 120	
QY	181 ggtctggagttccaagggaatctcaagccagtcctgtagaataatcttgataattctcg 240	
DB	121 ggtctggagttccaagggaatctcaagccagtcctgtagaataatcttgataattctcg 180	
QY	241 aaacttccgaagaagagctatacaacttcaagggtcaaccggagggaaccaagagca 300	
DB	181 aaacttccgaagaagagctatacaacttcaagggtcaaccggagggaaccaagagca 240	
QY	301 gtgcttcaagctattctgataaactgtgaaggtcaagaagaagcgagacgttccacagc 360	
DB	241 gtgcttcaagctattctgataaactgtgaaggtcaagaagaagcgagacgttccacagc 300	

[illegible]

RESULT	7
AAA96575	
ID	AAA96575 standard; DNA; 1221 BP.
XX	

AC	AA96575,	
XX		
DT	08-FEB-2001	(first entry)
XX		
DE	Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNA.	
XX		
KW	Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;	
KW	cancer; cardiovascular disorder; inflammatory disorder; asthma;	
KW	rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;	
KW	septic shock; adult respiratory distress syndrome; AIDS; cancer;	
KW	platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;	
KW	clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;	
KW	diverticulitis; ulcerative colitis; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	CA2296936-A1.	
XX		
PD	03-AUG-2000.	
XX		
PF	03-FEB-2000; 2000CA-2296936.	
XX		
PR	03-FEB-1999; 99US-0118674.	
XX		
PA	(GLYC-) GLYCDESIGN INC.	
XX		
PI	Korczak B, Lew A;	
DR	WPI; 2000-594746/57.	
XX		
PT	New nucleic acid molecules of core 2	
PT	beta-1,6-N-acetylglucosaminyltransferase useful for providing new	
PT	compositions for treatment of disorders mediated by the enzyme	
PT	including cancer, cardiovascular and inflammatory disorders.	
XX		
PS	Claim 4; Page 54; 66pp; English.	
XX		
CC	The present sequence encodes a fragment of a human core 2	
CC	beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.	
CC	The polypeptide can be used to treat diseases and disorders, such as	
CC	cancer, cardiovascular disorders and inflammatory disorders including	
CC	asthma, rheumatoid arthritis, inflammatory bowel disease,	
CC	arteriosclerosis, septic shock, adult respiratory distress syndrome	
CC	(ARDS) and cancer. Various platelet-mediated pathologies such as	
CC	atherosclerosis and clotting can also be treated. The polypeptides of	
CC	the invention are predominantly expressed in gastrointestinal tissue	
CC	(stomach, colon, intestine, testis) and are elevated in cancer.	
CC	Gastrointestinal disorders that may be prevented or treated include	
CC	ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis	
CC	and ulcerative colitis. The antibodies may be used in	
CC	immuno-histochemical analysis, to detect the novel polypeptide and to	
CC	localize it to particular cells and tissues and to specific subcellular	
CC	locations and to quantitate the level of expression.	
XX		
SQ	Sequence 1221 BP; 342 A; 285 C; 295 G; 299 T; 0 other;	
Query Match	64.8%; Score 1221; DB 21; Length 1221;	
Best Local Similarity	100.0%; Pred. NO. 0;	
Matches 1221: Conservative	0; Mismatches	0; Indels
		0; Gaps
Oy	157 ttgaagtgtgactctgcagcaactgtgtgtgagtcacaggaatctcaagcagactgt	216
Db	1 ttgaagtgtgactctgcagcaactgtgtgtgagtcacaggaatctcaagcagactgt	60
Oy	217 aggaatctctgtataatttcctggaactcttcagcaagaggtctatcaactgttcagg	276
Db	61 aggaatctctgtataatttcctggaactcttcagcaagaggtctatcaactgttcagg	120
Oy	277 gtaccacgaaggacacgaagcagtgcttgaagctatctcgaataactctggagtcagg	336
Db	121 gtaccacgaaggacacgaagcagtgcttgaagctatctcgaataactctggagtcagg	180

|||||
Db 781 gagtgaagaggagatgtgtgtagaactgtattcaattgtaattgcctctgttagctt 840
Oy 1707 ttccattctgtgagctgcgcttcccaataatccaggttggtagctgtagagaagact 1766
|||||
Db 841 ttccattctgtgagctgcgcttcccaataatccaggttggtagctgtagagaagact 900
Oy 1767 ttgattggaagaagaaaccttccctctgtactgttaacttaataataatagctcctgatt 1826
|||||
Db 901 ttgattggaagaagaaaccttccctctgtactgttaacttaataataatagctcctgatt 960
Oy 1827 caaagta 1833
|||||
Db 961 caaagta 967

RESULT 9
AAV59681
ID AAV59681 standard; DNA; 777 BP.
XX
AC AAV59681;
XX
DT 19-JAN-1999 (first entry)
XX
DE Human secreted protein gene 171 clone HTEDJ34.
XX
XX Human: secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; retinosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN MO9839448-r2.
PD
PD 11-SEP-1998.
XX
PF 06-MAR-1998: 98WO-US04493.
XX
PR 02-OCT-1997: 97US-0061060.
PR 07-MAR-1997: 97US-0038621.
PR 07-MAR-1997: 97US-0040161.
PR 07-MAR-1997: 97US-0040162.
PR 07-MAR-1997: 97US-0040163.
PR 07-MAR-1997: 97US-0040333.
PR 07-MAR-1997: 97US-0040334.
PR 07-MAR-1997: 97US-0040336.
PR 07-MAR-1997: 97US-0040626.
PR 11-APR-1997: 97US-0043311.
PR 11-APR-1997: 97US-0043312.
PR 11-APR-1997: 97US-0043313.
PR 11-APR-1997: 97US-0043314.
PR 11-APR-1997: 97US-0043358.
PR 11-APR-1997: 97US-0043569.
PR 11-APR-1997: 97US-0043576.
PR 11-APR-1997: 97US-0043578.
PR 11-APR-1997: 97US-0043580.
PR 11-APR-1997: 97US-0043669.
PR 11-APR-1997: 97US-0043670.
PR 11-APR-1997: 97US-0043671.
PR 11-APR-1997: 97US-0043672.
PR 11-APR-1997: 97US-0043674.
PR 23-MAY-1997: 97US-0047492.
PR 23-MAY-1997: 97US-0047500.
PR 23-MAY-1997: 97US-0047501.
PR 23-MAY-1997: 97US-0047502.
PR 23-MAY-1997: 97US-0047503.
PR 23-MAY-1997: 97US-0047581.
PR 23-MAY-1997: 97US-0047582.

PR 23-MAY-1997: 97US-0047583.
PR 23-MAY-1997: 97US-0047584.
PR 23-MAY-1997: 97US-0047585.
PR 23-MAY-1997: 97US-0047586.
PR 23-MAY-1997: 97US-0047587.
PR 23-MAY-1997: 97US-0047588.
PR 23-MAY-1997: 97US-0047589.
PR 23-MAY-1997: 97US-0047590.
PR 23-MAY-1997: 97US-0047592.
PR 23-MAY-1997: 97US-0047593.
PR 23-MAY-1997: 97US-0047594.
PR 23-MAY-1997: 97US-0047595.
PR 23-MAY-1997: 97US-0047596.
PR 23-MAY-1997: 97US-0047597.
PR 23-MAY-1997: 97US-0047598.
PR 23-MAY-1997: 97US-0047599.
PR 23-MAY-1997: 97US-0047600.
PR 23-MAY-1997: 97US-0047601.
PR 23-MAY-1997: 97US-0047612.
PR 23-MAY-1997: 97US-0047613.
PR 23-MAY-1997: 97US-0047614.
PR 23-MAY-1997: 97US-0047615.
PR 23-MAY-1997: 97US-0047617.
PR 23-MAY-1997: 97US-0047618.
PR 23-MAY-1997: 97US-0047632.
PR 23-MAY-1997: 97US-0047633.
PR 06-JUN-1997: 97US-0048964.
PR 06-JUN-1997: 97US-0048974.
PR 13-JUN-1997: 97US-0049610.
PR 08-JUL-1997: 97US-0051926.
PR 16-JUL-1997: 97US-0052874.
PR 18-AUG-1997: 97US-0053724.
PR 22-AUG-1997: 97US-0056630.
PR 22-AUG-1997: 97US-0056631.
PR 22-AUG-1997: 97US-0056632.
PR 22-AUG-1997: 97US-0056636.
PR 22-AUG-1997: 97US-0056637.
PR 22-AUG-1997: 97US-0056662.
PR 22-AUG-1997: 97US-0056664.
PR 22-AUG-1997: 97US-0056845.
PR 22-AUG-1997: 97US-0056862.
PR 22-AUG-1997: 97US-0056864.
PR 22-AUG-1997: 97US-0056872.
PR 22-AUG-1997: 97US-0056874.
PR 22-AUG-1997: 97US-0056875.
PR 22-AUG-1997: 97US-0056876.
PR 22-AUG-1997: 97US-0056877.
PR 22-AUG-1997: 97US-0056878.
PR 22-AUG-1997: 97US-0056879.
PR 22-AUG-1997: 97US-0056880.
PR 22-AUG-1997: 97US-0056881.
PR 22-AUG-1997: 97US-0056882.
PR 22-AUG-1997: 97US-0056884.
PR 22-AUG-1997: 97US-0056886.
PR 22-AUG-1997: 97US-0056887.
PR 22-AUG-1997: 97US-0056888.
PR 22-AUG-1997: 97US-0056889.
PR 22-AUG-1997: 97US-0056892.
PR 22-AUG-1997: 97US-0056893.
PR 22-AUG-1997: 97US-0056894.
PR 22-AUG-1997: 97US-0056903.
PR 22-AUG-1997: 97US-0056908.
PR 22-AUG-1997: 97US-0056909.
PR 22-AUG-1997: 97US-0056910.
PR 22-AUG-1997: 97US-0056911.
PR 05-SEP-1997: 97US-0057650.
PR 05-SEP-1997: 97US-0057669.
PR 05-SEP-1997: 97US-0057761.
PR 12-SEP-1997: 97US-0058785.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI

PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni Y, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX WPI: 1998-506364/43.
DR P-PSDB: AAW74898.
XX

PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX

Claim 1; Page 410-411; 721pp; English.

XX This sequence represents a nucleic acid molecule designated Gene 171
XX from the human cDNA clone HTEB134 (deposited as clone ATCC 97904 and
XX ATCC 209050) which encodes a secreted human protein. The gene can be
XX used to generate fusion proteins by linking to the gene to a human
XX immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
XX the fused protein as compared to the human protein only.
XX The invention relates to 186 novel genes and their fragments (nucleic
XX acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological
XX conditions can be diagnosed by determining the amount of the new
XX polypeptides in a sample or by determining the presence of mutations in
XX the new polynucleotides. Specific uses are described for each of the 186
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AAV59511 for described uses).

SQ Sequence 777 BP; 203 A; 142 C; 194 G; 198 T; 40 other;

Query Match 25.8%; Score 486.2; DB 19; Length 777;
Best Local Similarity 81.9%; Pred. No. 4.1e-146;

Matches 610; Conservative 34; Mismatches 67; Indels 34; Gaps 7;

QY 1123 tcagacatgactctatgtccagcagctgtgcaagtgccaggttcattgagagacatcgat 1182
Db 8 kcagacatgactctatgtccagcagctgtgcaagtgccaggttcattgagagacatcgat 67

QY 1183 aagggtgctccttctgctcctgctctggaatccacagcggtctatcggttattgg 1242
Db 68 aagggtgctccttctgctcctgctctggaatccacagcggtctatcggttattgg 127

QY 1243 gctgggagcttgaaatgtgacttcaaacacatccgctgtggccaaagatttacc 1302
Db 128 gctgggagcttgaaatgtgacttcaaacacatccgctgtggccaaagatttacc 187

QY 1303 aaggtagatgaatgtccttcagtgcttagaagatacactacgtataaagcc---at 1358
Db 188 aaggtagatgaatgtccttcagtgcttagaagatacactacgtataaagccatctc 247

QY 1359 ctatggagctgaactttgagacacatc-----gagagcgttgctacc 1402
Db 248 atggagacctgaactttgagacacatcgaagagcgcttgcttcgctgttgaggc 307

QY 1403 tgtgggcaagagctgacgaacatgctcgaacttggagac---agtgtgggtgg 1460
Db 308 caagagagatgttccacaacatgtyccargamtytgykxggamcarakkkkxgggga 367

QY 1461 gaccagggccttgcaa---ttcgtggcactctta--ggataagaagggcgtctataat 1515
Db 368 rrmcmrggtytctscaanvctskkgcmwccytlttagarrgrrggcgkctwtatagat 427

QY 1516 tgtgggta---agtagctcttggccttgcaaatgtcgtcgtgggtgaagctcgt 1571
Db 428 tgtgggtaagtagatcttgcctctgcaaatgtcgtcgtgggtgaagctcgt 487

QY 1572 tctcc---acccttaaccctagtagtccctccactaacttctccactaagtgaaagga 1628
Db 488 tctctctmaccccttaaccctagtagtccctccactaacttctccactaagtgaaagga 547
QY 1629 gaactgctgtataggagagtgaaagggatagtggttagagcacttgatttcagttg 1688

Db 548 gaactgctgtataggagagtgaaagggatagtggttagagcacttgatttcagttg 607
QY 1689 aatgctgtgttagcttccatctctgtgagcgtccgtccctataatccagttg 1748
Db 608 atgctgtgttagcttccatctctgtgagcgtccgtccctataatccagttg 667
QY 1749 gttagctgtgagagactttagtgaagagaccttccctctgtactgttaactaa 1808
Db 668 gttagctgtgagagactttagtgaagagaccttccctctgtactgttaactaa 727
QY 1809 aataatagctctgattcaagta 1833
Db 728 aataatagctctgattcaagta 752

RESULT 10

AAA96571
ID AAA96571 standard; DNA; 524 BP.

AAA96571;
XX 08-FEB-2001 (first entry)

XX Noncoding region of core 2 beta-1,6-N-acetylglucosaminyltransferase gene.

KW Core 2 beta-1,6-N-acetylglucosaminyltransferase: core2b GlcNAc-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.

XX Homo sapiens.

XX CA2296936-A1.

XX 03-AUG-2000.

XX 03-FEB-2000; 2000CA-2296936.

XX 03-FEB-1999; 99US-0118674.

XX (GLYC-) GLYCDESIGN INC.

XX Korcek B, Lew A;

XX WPI: 2000-594746/57.

PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.

PS Disclosure; Page 52; 66pp; English.

XX The present sequence represents a noncoding region from a human core 2
XX beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) gene.
XX The polypeptide can be used to treat diseases and disorders, such as
XX cancer, cardiovascular disorders and inflammatory disorders including
XX asthma, rheumatoid arthritis, inflammatory bowel disease,
XX arteriosclerosis, septic shock, adult respiratory distress syndrome
XX (ARDS) and cancer. Various platelet-mediated pathologies such as
XX atherosclerosis and clotting can also be treated. The polypeptides of
XX the invention are predominantly expressed in gastrointestinal tissue
XX (stomach, colon, intestine, testis) and are elevated in cancer.
XX Gastrointestinal disorders that may be prevented or treated include
XX ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
XX and ulcerative colitis. The antibodies may be used in
XX immuno-histochemical analysis, to detect the novel polypeptide and to
XX localize it to particular cells and tissues and to specific subcellular
XX locations and to quantitate the level of expression.

XX Sequence 524 BP; 171 A; 90 C; 127 G; 136 T; 0 other;

Query Match 24.2%; Score 456; DB 21; Length 524;
Best Local Similarity 100.0%; Pred No. 1.8e-136;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 gacacactatgagagctgtgtctactctgtgtgggaagagacatgtacaacaatgtctcgaac 1437
DB 1 gacacactatgagagctgtgtctactctgtgtgggaagagacatgtacaacaatgtctcgaac 60
QY 1438 ttgtctggagacagtgtgtgtgggaagacaggtcttgcactgtgtgcactcttaggata 1497
DB 61 ttgtctggagacagtgtgtgtgggaagacaggtcttgcactgtgtgcactcttaggata 120
QY 1498 agagggctgtctatagatttggtaagtatgactttgtccttgcgaattgtcgtcggg 1557
DB 121 agagggctgtctatagatttggtaagtatgactttgtccttgcgaattgtcgtcggg 180
QY 1558 tgaatgtctgttcttctcacccttaacctagtagttcctcactaacttctcacta 1617
DB 181 tgaatgtctgttcttctcacccttaacctagtagttcctcactaacttctcacta 240
QY 1618 agtgaagaatgagacgtctgtatagaggagaagtgaaggagatatgtgttagagcactt 1677
DB 241 agtgaagaatgagacgtctgtatagaggagaagtgaaggagatatgtgttagagcactt 300
QY 1678 gatttcagttgaaatgtcgtgtgtgttagctttcattctgtggagcgtgcgttccataaa 1737
DB 301 gatttcagttgaaatgtcgtgtgtgttagctttcattctgtggagcgtgcgttccataaa 360
QY 1738 ttccaggtttgttagagctgtgtagaagagaacttgaatgaaagaacacttccctctgact 1797
DB 361 ttccaggtttgttagagctgtgtagaagagaacttgaatgaaagaacacttccctctgact 420
QY 1798 gtttaacttaataatagctccttgattcaagta 1833
DB 421 gtttaacttaataatagctccttgattcaagta 456

RESULT 11

AAx24043 standard; cDNA; 2204 BP.

AAx24043;

28-JUN-1999 (first entry)

Human core 2Gnt DNA.

Screening; treatment; prevention; cardiomyopathy; inhibitor;
diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
UDP-GlcNAc:Galbeta1-3GalNAc-alphaR; transgenic animal; germ line;
beta1,6-N-acetylgalucosaminyl-transferase; human; core 2 Gnt; ss.

Homo sapiens.

Location/Qualifiers

FT CDS 244..1530

FT /tag a

CA2186987-A.

02-APR-1998.

02-OCT-1996; 96CA-2186987.

02-OCT-1996; 96CA-2186987.

(MOUN) MOUNT SINAI HOSPITAL CORP.

PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

DR WPI; 1998-399608/35.

DR P-PSDB; AAW93943.

PT Screening for substances that prevent or treat cardiomyopathy
associated with diabetes and hyperglycaemia - comprised reacting
core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
in presence of test substance

PS Disclosure; Fig 9; 35pp; English.

CC This invention describes a method for screening for a substance that
CC may be used to prevent or treat cardiomyopathy associated with diabetes
CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
CC (UDP-GlcNAc:Galbeta1-3GalNAc-alphaR beta-1,6-N-acetylgalucosaminyl-
CC transferase) with an acceptor substrate and a sugar nucleotide donor in
CC the presence of a test substance under conditions whereby the core 2
CC GlcNAc-T produces a reaction product, determining the amount of reaction
CC product, and comparing the amount of reaction product with the amount
CC obtained in the absence of the test substance, where lower amounts of
CC reaction product in the presence of the test substance indicate that the
CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
CC methods for preventing or treating cardiomyopathy associated with
CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
CC germ cells and somatic cells all contain a DNA construct introduced into
CC the animal or an ancestor of the animal at an embryonic stage, where
CC incorporation of the DNA construct into the germ line of the animal
CC causes the animal to develop cardiomyopathy similar to that associated
CC with diabetes mellitus and hyperglycaemia. This sequence encodes the
CC human core 2 Gnt protein (beta-1,6-N-acetylgalucosaminyltransferase)
CC which is used to describe the method of the invention.

SO Sequence 2204 BP; 641 A; 414 C; 498 G; 651 T; 0 other;

Query Match 19.6%; Score 369.4; DB 19; Length 2204;
Best Local Similarity 60.2%; Pred. No. 4.4e-108;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 337 aagaagcgagagcctttcacagacaccacactactcctccctcacagagactgtggac 396
DB 490 aagaagcgccctcgtgtggaacactcctgactatataaacatgcccgtgacttcttcc 549
QY 397 ttcaagctgaaagagtcatcacagttccacatgacaagaagagtgtgtccct 456
DB 550 ttcatcaagagcgaataatattgtagaacccttagaagaagaagaggtgttcca 609
QY 457 atgcatactctatgtgtatcatgtagaagattgaaactttgaagctcagagct 516
DB 610 atgcatactctatgtgtatcatgtagaagattgaaactttgtacagagctcgtgaagc 669
QY 517 gtgtagccctcagaacataactgtgtccatgtgtgtagatgagaagctccccaacttcc 576
DB 670 atctatatgctcagaattctctatgtgtcattgtagaaccaaatccagagattcctat 729
QY 577 aagaagcggtcaaaagcaattatttctgtctcccaatgtcttccatagccagtaagctg 636
DB 730 tttagcgaagtgatggtcgcctgtctgttttagtaagtcttctgtgcccagcgatg 789
QY 637 gtccgggtgtttatctcctcctgtgcacgggtgacaagtacccaactgtatgaaagc 696
DB 790 gagagtggtttatgcatgtgagcgggttcaagctacccaactgtacatgaaagat 849
QY 697 ttgtccagagctcagtgccgtgtgaaatacttctcgaatacagtgtgagcgacttctc 756
DB 850 ctctatgcaatgagtgaactgtgaagtactgtataacttgtgtatgtgttcc 909
QY 757 ataaagcaatgagagatgtgtccaggtcccaagatgtgatatgtggaataagatg 816
DB 910 attaaacccaactagaatgttgcaggaagctaaagtgtttatgtggaagaaacaacgtg 969
QY 817 ggtcagaggtactcctcctaagcacaagaaccgctggaatacatcattgtggtagt 876

Db 970 gaaacgagagatgcacatcccatataaagaagaagtgtagaagcggtatgagtcgtt 1029
Qy 877 agagacattacacacttaacacaaagaagaagatcctcccttaattactatg 936
Db 1030 aatggaagcgtgacaacaagacgtgtcaaatgtctccccaactcgaaacactctc 1089
Qy 937 ttacaggaatgctgtacatgtgtctcccgagatcttcgtcccaacatgctttgaagac 996
Db 1090 ttcttgccagtgctactcttgctgtcagtagagatgtggtgggtactactaagat 1149
Qy 997 cctaattcccaacactgtatggaatgggtaaaagacactttagccagatgaacactc 1056
Db 1150 gaaaataaccacaaagttagtgagtgacacagacacatacagccctgagatctc 1209
Qy 1057 tgggcccacttcagcgcgacggttgatgctgtctgttcccaaccacccaagac 1116
Db 1210 tgggcccacacacaaagatccctgaagtcctcgaggtccactcccgacagcataagtat 1269
Qy 1117 gacatctcagacatgactctatctatgtccagcgtgtgcaagtgtcagtgaggaagac 1176
Db 1270 gatcatctgacatgcaagacagtgctgacaggtttgtcgaagtgtgactttgaggggtgat 1329
Qy 1177 atcgataaggtgtgtctccttaatgtccctcgtctgtgaatccacagcggtctatctgctt 1236
Db 1330 gttcccaagggtgtctccctaccccgctcgatgtagtgcatagtgtcgtcagtgatcatt 1389
Qy 1237 fatgagggctgggagcttgatgtatgttcaaaacataccctgctgttgcaacaagttt 1296
Db 1330 ttcggagctgtgactgtgaactgtgactgtgcgaacacacactgtgttgcaataaagttt 1449
Qy 1297 gaccacaagtagatgataatgctctcctcagtgctgaagaataacatacgtatataaggc 1355
Db 1450 gacgtgagtgatgtgacctcttgccatccagtggttgatgagcattttagacacaaagc 1508

RESULT 12

AAC65469 ID AAC65469 standard; cDNA; 2204 BP.

AC AAC65469;

DT 12-FEB-2001 (first entry)

DE Human heart core 2 GlcNAc-T coding sequence.

XX Human: diabetes: core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;

KW UDP-GlcNAc:Galbeta1-3GalNAcalphaPhar beta1-6

KW N-acetylglucosaminyltransferase; ss.

OS Homo sapiens.

PN US6131578-A.

PD 17-OCT-2000.

PE 02-OCT-1997; 97US-0943058.

PR 02-OCT-1996; 96US-0046876.

PA (KING/) KING G L.

PA (NISH/) NISHIO Y.

PA (KOYA/) KOYA D.

PA (DENN/) DENNIS J W.

PA (WAR/) WARREN C E.

PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
DR WPI: 2000-678642/66.
DR P-PSDB; AAB30298.
XX
PT Preventing or treating cardiomyopathy associated with diabetes mellitus
and hyperglycemia, comprises administering a substance that inhibits

PT core 2 N-acetylglucosamine-T activity -
PS Claim 3; Fig 9; 21pp; English.
XX The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc
CC alphaPhar beta1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
CC and coding sequences. The enzyme is associated with cardiomyopathy in
CC diabetes and hyperglycemia sufferers. The invention also provides
CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
CC in treatment.
XX
SQ Sequence 2204 BP; 641 A; 414 C; 498 G; 651 T; 0 other;
Query Match 19.6%; Score 369.4; DB 21; Length 2204;
Best Local Similarity 60.2%; Pred. No. 4,4e-108;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;
Qy 337 aagaagcgagagccttccacagacacccactactctccctccacagagactgtgagac 396
Db 490 aaaaagcgccctcggttgacacacctgacgactataataacatgaccagtgcttctct 549
Qy 337 ttcaagcgtgaaggaagtcatatacagttccactgagcaagaagagtggtccct 456
Db 550 ttccaaagagacgcaaatatattgtgaaccccttagtaagaagagcgagtttcca 609
Qy 457 attgatactatgtagtattcatatgagaagattgaacattgaagagctactcgagct 516
Db 610 atagatatctctatagtggttcatcacaagaattgaagcttgtagagctgtcgagggc 669
Qy 517 gtgtatgccctccacaaataactgtgtccatgtggaatggaaggtcccgcaaaacttc 576
Db 670 atcatatgctccagaattctcatgtgtcatgtgagacacaaatccgagattccat 729
Qy 577 aaagagcggtgcaaaagcaatattcttctgtctcccaatgttccatagccagtgactg 636
Db 730 ttagctgagtgatgtggtgactgcttccgttttagtaatgtcttgtgtgacgagcgatg 789
Qy 637 gttcggtgtgtttatgctcctcgtgcaggggtgcaagctgacactgaactgaggaagac 696
Db 790 gagggtgtgtttatgcatctgtgagcggttcacagctgacactgaactgaggaagat 849
Qy 697 ttgtcccaagctcagtgccgttggaataactcctgaaatactgtggaacggaacttct 756
Db 850 ctctatgcaatgagtgcaaacactggaagtactgtgaaatcttftgtgtagtattcc 909
Qy 757 ataaagcaatgacagagaatgttccaggtctcctaagatgttgaatgaggaatgactgt 816
Db 910 attaaacccaactagaattgttcaggaagctcaagttgttaatgtggaataacacactg 969
Qy 817 ggtcagaggtaccccttaagcacaagaaccccgctggaatatcatcttggagtgagt 876
Db 970 gaaacggaagagatgcatcccatataaagaagaagatggaagaacggtatgagtgctt 1029
Qy 877 agagacattacacacttaacacaaagaagaagatccctcccttaattactatg 936
Db 1030 aatggaagcgtgacaacaacagggagctgtcaaatgtctccccaactcgaaacactctc 1089
Qy 937 ttacaggaatgctgtacatgtgtctcccgagatcttcgtcccaacatgctttgaagac 996
Db 1090 ttcttgccagtgctactcttgctgtcagtagaggtatgtggtggtactactaagat 1149
Qy 997 cctaattcccaacactgtatggaatgggtaaaagacactttagccagatgaacactc 1056
Db 1150 gaaaataaccacaaagttagtgagtgacacagacacatacagccctgagatctc 1209
Qy 1057 tgggcccacttcagcgcgacggttgatgctgtctgttcccaaccacccaagac 1116
Db 1210 tgggcccacacacaaagatccctgaagtcctcgaggtccactcccgacagcataagtat 1269
Qy 1117 gacatctcagacatgactctatctatgtccagcgtgtgcaagtgtcagtgaggaagac 1176
Db 1270 gatcatctgacatgcaagacagtgctgacaggtttgtcgaagtgtgactttgaggggtgat 1329

Oy	1177	atcgataaggtgtcctcttatgtctccctgtctctcgataccacacgcggcgtctcgtgt	1235		
Db	1330	gtttccaaagggtgtccctccctaccgccctcgatgagtcacatgtgcgtccagtgtgcat	1388		
Oy	1237	tatggagctggggacattgaaattgatgcttcaaacaccatcacctgtttgccaacagatt	1296		
Db	1330	ttcggagctgggtgtcattgacattgagatgcgcgcgaacacacactgtttgcacataagtt	1449		
Oy	1297	gaccacaaggtagatgataatgtactcttcacgtcttgaagaatlaactaagttataagc	1355		
Db	1450	gacgtgagatgtgtagcctcttgcacatccagtggtttgattgagcatattggacacaaagc	1508		
RESULT 13					
ID	AA061559	standard; cDNA; 2105 BP.			
XX	XX				
AC	AA061559;				
XX	XX				
DT	07-OCT-1994	(first entry)			
XX	XX				
DE	cDNA sequence of human core 2 beta 1-6				
DE	N-acetylglucosaminyltransferase (C2Gnt or core 1-6 AGT).				
XX	XX				
KW	C2Gnt: 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;				
XX	O-glycan; ss.				
OS	Homo sapiens.				
XX	XX				
Key	Location/Qualifiers				
FT	CDS	220..1504			
FT		/*tag= a			
FT	polyA_signal	1913..1918			
FT		/*tag= b			
FT	misc_signal	248..314			
FT		/*tag= c			
FT		/label= signal/membrane-anchoring domain			
XX	XX				
PM	EP930747-A.				
XX	XX				
PD	06-APR-1994.				
XX	XX				
PF	29-SEP-1993;	93EP-0250268.			
XX	XX				
PR	01-OCT-1992;	92US-0955041.			
XX	XX				
PA	(LJOL-) LA JOLIA CANCER RES FOUND.				
XX	XX				
PI	Bierhuizen MFA, Fukuda M;				
XX	XX				
DR	WPT: 1994-111195/14.				
XX	XX				
DR	P-PSDB; AAR51386.				
XX	XX				
PT	New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are				
PT	used for the study of the effect of variant O-glycan(s) on				
PT	cell-cell interactions, partic. in cancers				
XX	XX				
PS	Disclosure; Page 20-22; 34pp; English.				
XX	XX				
CC	C2Gnt, or an active fragment thereof, catalyses the formation of				
CC	critical branches in O-glycans. cDNAs encoding various				
CC	glycosyltransferases can be isolated by transient expression of cDNA				
CC	in recipient cells, e.g. COS-1. COS-1 cells were transfected				
CC	with a cDNA library, pCDR alpha-2P1, constructed from poly(A)+ RNA				
CC	of activated T lymphocytes which express the C2Gnt. Transfected				
CC	cells were selected using Mab T305, which identifies a				
CC	hexasaccharide on leukosialin. Leukosialin CD43 is an acceptor				
CC	molecule for C2Gnt activity. A plasmid, pCDR alpha-lieu,				
CC	which directed expression of the T305 antigen was identified. The				
CC	cDNA insert was isolated and sequenced. The cDNA encoded the				
CC	acceptor molecule leukosialin CD43.				
XX	XX				

Seq	Sequence	2105 BP: 614 A: 402 C: 500 G: 589 T: 0 other:
Query Match	19.5%;	Score 367.8; DB 15; Length 2105;
Best Local Similarity	60.1%;	Pred. No. 1.4e-107;
Matches 612: Conservative	0;	Mismatches 407; Indels 0; Gaps 0;
QY	337 aagaagcgagagccttccagagaccaccatctctccctaccagagactgtaagac	396
DB	466 aaaaagcgccctcgglygacacctgacgactatataaataccagatggttctt	525
QY	397 ttcaagcgctgaagaagagttcataagttccactgagcagaagaaggtggagttcc	456
DB	526 ttcatcaagagagcccaataatattgtagaacctcttagtaagaagaagcgagttcca	585
QY	457 attgcatactatggttgatcaatgagaagaattgaaaactttgaagctactcgagct	516
DB	586 atagataattcatagttggttcatcaacaagaattgtaaatctgtgaacagctgtgagttgc	645
QY	517 ggtgtccccccaagaacatactacgtgtgcatctgtgagtaggaagttccccgaaccttc	576
DB	646 actataatccccgaattccatctcatgctgtctcgtgtgacaacaataatccgagattccat	705
QY	577 aagaagcgcgltcaagaacattattcttcgtctcccaaatgltctcataagccgtaagctg	636
DB	706 tttagctgcagtagatgggcacgtcgtctcgttttagtaatgctttgtgtgcagccgattg	765
QY	637 gttcgggtgtgttaagccctcgtgcccaggggtgcagagctgacttaactgcgttggaagc	696
DB	766 gagaatgtgttcatagcatcgtcgagccgggtlccaggtccgaactcaacgtcatgaagat	825
QY	697 ttgcctccagagctcagttgcgtgagaaataactcttcggaatcacatgtggtggagacttcc	756
DB	826 ctctatgcagtagatacaaacctggaagtagtactgtataaactcttgtcgtatgatttccc	885
QY	757 ataaagagcgaatgcgcgagagttgtlccagctctcccaagattgtgaatgggaggaatgcag	816
DB	886 attaaaaccacactgaactgttgcaggaaactccaagttgttaatggtggaataaacactcg	945
QY	817 ggtcagagatgccctcctcctaagcacaagaacccgttggaatatcatctttagtagatg	876
DB	946 gaaacggagagagatgcatccatccataaagaagaagaaggttggaagaagcgtatgagttcgt	1005
QY	877 agagacacattacacctcaacccaagaagaagatcctccccttaataatttaactaig	936
DB	1006 aatggaagactgacaacaaccagagagctgcacaatagtctccctccacactcgaaacacttcc	1065
QY	937 tttaacagggaatgcgtacatttggtcttcccggaatttgcgtcaacatttttggaaac	996
DB	1066 ttcttcgtgagtgctactctgttgcgttcagtagggagatggtggattgtactacagaat	1125
QY	997 cctaaatcccaacaactgtattgaatggtgtaaagaacactataagccccaatgaacactc	1056
DB	1126 gaaaataatccaaaagttagtgtagtgggcacagaacacatacagccctatagatattac	1185
QY	1057 ttgggacaccccttcagcggtgcaggtgtagtggtgctgtctgttcccaacccaacccaagtc	1116
DB	1186 ttgggccaacacccaagaagatctccgtgaagctccggggtccctccctgcgcagccataagat	1245
QY	1117 gacatctcagaaatactctatcttccaggtcgcgttgcgaagtggcaggtlcatgaggaagc	1176
DB	1246 gattatctcgaatgcaagcaggttgcaggtttgttcaagtgtgcagttactttagaggtgat	1305
QY	1177 atcgaataaggtgtcctcttaattgtccctgctctctgtgaatccacacagcggtctatctgcgt	1236
DB	1306 gtttcccaaggggtgtccctaccccgcccttgcgttagagttcatgtgcgtcgaattgtgcatt	1365
QY	1237 tatgggctgcgggagctgaattgagttggtcttcaaaacacactgcgtttggccaacaagtt	1296
DB	1366 ttccgagctcgtggaacttgaactgagatgctgtgcgcaaacacacacttgtttggccaataagtt	1425
QY	1297 gaccacaagttatgatatatgtcctctcaatggtcttagaagaatatcactacgttataagc	1355

Db 1426 gacgtgatgttgacctcttggccatccagtggttgatgagcatltgagacacaagc 1484

RESULT 14

ID AAX24042 standard; cDNA; 5010 BP.

AC AAX24042;

DT 28-JUN-1999 (first entry)

Rat DH1 cDNA.

KW DH1; ra; screening; treatment; prevention; cardiomyopathy; inhibitor;
 KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
 KW UDP-GlcNAc:Galbeta1-3GalNAc-alpha; transgenic animal; germ line;
 KW beta-1,6-N-acetylglucosaminyl-transferase; ss.

OS Rattus sp.

FH	Key	Location/Qualifiers
00		000000

/*tag= a

[illegible]

XX

XX
XX

XX

XX
XX
XX

XX

XX
XX
1000
1000
1000
1000
1000
1000

DR P-PSDB; AAW93942.

PT Screening for sub

PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
in presence of test substance
XX
PS Example 1; Fig 3A; 35pp; English.

CC This invention describes a method

CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
CC (UDP-GlcNAc:Galbeta1-3GalNAc-alphaR beta-1,6-N-acetylglucosaminyl-
CC transferase) with an acceptor substrate and a sugar nucleotide donor in
CC the presence of a test substance under conditions whereby the core 2
CC GlcNAc-T produces a reaction product, determining the amount of reaction
CC product, and comparing the amount of reaction product with the amount
CC obtained in the absence of the test substance, where lower amounts of
CC reaction product in the presence of the test substance indicate that the
CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
CC methods for preventing or treating cardiomyopathy associated with
CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
CC germ cells and somatic cells all contain a DNA construct introduced into
CC the animal or an ancestor of the animal at an embryonic stage, where
CC incorporation of the DNA construct into the germ line of the animal
CC causes the animal to develop cardiomyopathy similar to that associated
CC with diabetes mellitus and hyperglycaemia. This sequence encodes the rat
CC DH1 protein which is used to describe the method of the invention.
XX
XX Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match

Matches	659;	Conservative	0;	Mismatches	469;	Indels	3;	Gaps	1;
---------	------	--------------	----	------------	------	--------	----	------	----

QY	251	caagaagctctatcaactctgtccagggtccaccggaagggaacaaagacagtgctcaag	310
Db	959	ccaataagaatgtaattcaatcccaagaagttttacagaggtgacccgaagaatattccgaag	1018
QY	311	ctattctgtaaacactgtaggttcaa---ggaagaagcgaaagcctttccaaagacacacac	367
Db	1019	tgaagctctgagataactaacgctgcaatttcaagaagcgctccggcggaacccgcatgct	1078
QY	368	accctccctaccagaagactgtgagacattcaaaagcgtgaaggaagttcataacgcttc	427
Db	1079	atataaacaatgacccggaagactgcgcctcccttcaacgacagcaaatataatgtaggc	1138
QY	438	cactggaacaaagaaaggtgtagtctccctatcaacactatgtgtatctatgaa	487
Db	1139	cccttactcaagaagaaaggtgtgcttccaatgtcatcttcatagtgtttcaactaa	1198
QY	488	ctgaaacattgtaaaaggctactcgcgaagctgtgtatgccccctcagaacataactgtgc	547
Db	1199	ttgacatctgtgaacagcctcctctagagccactatatactgcagaagattctactgtatc	1258
QY	548	atgtggaagagaagtcgcccaagaacttcaagaagcggtcaaaacaaatattcttgc	607
Db	1259	acgtggaacagaagaagcagagaatcctttttaagcgggtgcaaggtattgcaaccgct	1318
QY	608	tcccaaaatgtcttcaatagcaacagtaagcctgtgttcgggttttatgtctcctgttccaa	667
Db	1319	ttgtataatgtcttgttgccagccagctgtgagaggtgtgtatagcctccctcgtgtagc	1378
QY	668	tgcagaactgacactcaactgcatagtgaagaactgtctccagaagccgaagtcggtagaaatc	727
Db	1379	ttaaagctgcagactcaactgtcatgtatgaagacccgttatagaatgtaatgcaaatgtgaag	1438
QY	728	tcctgtaatacactgtgggaacgacttccctctaaagaagcaatgagatgtgtccagctc	787
Db	1439	tgtatcaatctttgtgtgatgtatgtattccctattaaacacacactgtgaatttgcagaagc	1498
QY	788	tcaaagatgttgaaltgvgaggaataagcatgtgagtcagaagtlacctccctaagcaaaaga	847
Db	1499	tcaagctcttccacaggggaaacaagccgtgaaactgtagaagatgtcctcccaacaagaag	1558
QY	848	ccgcgttcgaataatcactcttggtagttagtggagacatattacaactcaacaacaagaag	907
Db	1559	agaggttvgaaaaaaacgatacagcgtctgtgtgacgggaagcctgcgaacaaacatvgagtca	1618
QY	908	aggatccctcccccctataatctaactatgatttatacagaagatgcgtacattgtgcttcc	967
Db	1619	aggcgcagcgcctccacataaacctccctcttccagcagcgcctatttgcgtgcatcta	1678
QY	968	gagatttcgtccaaacatgcttttgaaagaacccataatcccaacaactgtattgagttaa	1027
Db	1679	ggagataatgtagtgatgtatgtctgtgaaataaaaataatctaaagttcatgaaatgtggac	1738
QY	1028	aagacactatagaacgcagatgaaacacactcgtggccaccccttaagcgtgacggtgagtc	1087
Db	1739	aggaacataataagcccgagatgagttctcctcgtggccacatcccaagaagatccctgaagtc	1788
QY	1088	ctggcctcgttcccaaccaccccaagtaagacatctcagacatgacattctattgccagc	1147
Db	1799	ctggtcttccctcccaagcacaataagtatattgtctcattgataatgtagtgcgtaggt	1858
QY	1148	tgtgcaagtgtgacaggtgcataagggagacatcgataaggtgtcctctatgtctccgtc	1207
Db	1859	tgtgcaagtgtgcaatattcgaagaagcgatgttccaatgvcgctccattatccacccgtgca	1918
QY	1208	ctggaaatccacaaagcgagcatcgtcgttatgtgggtctgggagattgtgaatgtgcttc	1267
Db	1919	gtggagatccatgtgcgtctgtgtgctgtcttggagttgtgagcttgagcgtgagatgcgc	1978
QY	1268	aaaacatcacctgttggccaacaagaagttgacccaagaagtaataatgtcttcag	1327
Db	1979	gcaaacacacatttcttggccaataaagttatgacatgtagatgtgattcccttgcctccag	2038

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:25:28 ; Search time 94.91 Seconds
(without alignments)
4875.920 Million cell updates/sec

Title: US-09-874-390-1_COPY_436_2319
Perfect score: 1884
Sequence: 1 gatgtgctctccaccct.....ataataatcaacaga 1884

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_MN:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1798.8	95.5	2128	3	US-09-233-506-1 Sequence 1, Appl1
2	369.4	19.6	2105	1	US-07-955-041-3 Sequence 3, Appl1
3	369.4	19.6	2105	1	US-08-227-455-3 Sequence 3, Appl1
4	369.4	19.6	2105	1	US-08-472-482-3 Sequence 3, Appl1
5	369.4	19.6	2105	1	US-08-487-069-3 Sequence 3, Appl1
6	355	18.8	2102	3	US-09-063-237-3 Sequence 3, Appl1
7	180.4	9.6	1807	1	US-08-118-906-13 Sequence 13, Appl1
8	180.4	9.6	1807	1	US-08-486-196-13 Sequence 13, Appl1
9	180.4	9.6	1807	1	US-08-488-135-13 Sequence 13, Appl1
10	180.4	9.6	1807	2	US-08-474-065-13 Sequence 13, Appl1
11	162.8	8.6	378	1	US-08-118-906-3 Sequence 3, Appl1
12	162.8	8.6	378	1	US-08-486-196-3 Sequence 3, Appl1
13	162.8	8.6	378	1	US-08-488-135-3 Sequence 3, Appl1
14	162.8	8.6	378	2	US-08-474-065-3 Sequence 3, Appl1
15	134	7.1	378	1	US-08-118-906-1 Sequence 1, Appl1
16	134	7.1	378	1	US-08-486-196-1 Sequence 1, Appl1
17	134	7.1	378	1	US-08-488-135-1 Sequence 1, Appl1
18	134	7.1	378	2	US-08-474-065-1 Sequence 1, Appl1
19	129.6	6.9	192	3	US-09-233-506-9 Sequence 9, Appl1
20	84	4.5	147	3	US-09-233-506-13 Sequence 13, Appl1
21	43.6	2.3	66	1	US-08-118-906-11 Sequence 11, Appl1
22	43.6	2.3	66	1	US-08-486-196-11 Sequence 11, Appl1
23	43.6	2.3	66	1	US-08-488-135-11 Sequence 11, Appl1
24	43.6	2.3	66	2	US-08-474-065-11 Sequence 11, Appl1
25	43	2.3	99	1	US-08-118-906-7 Sequence 7, Appl1
26	43	2.3	99	1	US-08-486-196-7 Sequence 7, Appl1
27	43	2.3	99	1	US-08-488-135-7 Sequence 7, Appl1

28	43	2.3	99	2	US-08-474-065-7 Sequence 7, Appl1
29	35	1.9	99	1	US-08-118-906-5 Sequence 5, Appl1
30	35	1.9	99	1	US-08-486-196-5 Sequence 5, Appl1
31	35	1.9	99	1	US-08-488-135-5 Sequence 5, Appl1
32	35	1.9	99	2	US-08-474-065-5 Sequence 5, Appl1
33	33.2	1.8	1941	4	US-09-402-002-1 Sequence 1, Appl1
34	33	1.8	11283	2	US-08-603-753D-3 Sequence 3, Appl1
35	33	1.8	11283	3	US-09-099-753-3 Sequence 3, Appl1
36	33	1.8	11283	4	US-08-986-106-3 Sequence 3, Appl1
37	32.8	1.7	246240	2	US-08-724-394A-20 Sequence 20, Appl1
38	32.8	1.7	246240	2	US-08-724-394A-21 Sequence 21, Appl1
39	32.8	1.7	246240	2	US-08-724-394A-22 Sequence 22, Appl1
40	32.2	1.7	3164	1	US-07-723-002C-3 Sequence 3, Appl1
41	32.2	1.7	7218	1	US-08-232-463-14 Sequence 14, Appl1
42	31.8	1.7	900	1	US-08-218-026-1 Sequence 1, Appl1
43	31.8	1.7	900	2	US-08-653-632-1 Sequence 1, Appl1
44	31.8	1.7	1653	1	US-08-218-026-3 Sequence 3, Appl1
45	31.8	1.7	1653	2	US-08-653-632-3 Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-233-506-1
; Sequence 1, Application US/09233506
; Patent No. 6136580
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jiumn-Chern
; TITLE OF INVENTION: A Beta1-1-6-N-Acetylglucosaminyltransferase That Forms
; TITLE OF INVENTION: Core 2, Core 4 and I Branches
; FILE REFERENCE: P-IJ 3415
; CURRENT APPLICATION NUMBER: US/09/233, 506
; CURRENT FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2128
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (354)..(1670)
US-09-233-506-1

Query Match 95.5%; Score 1798.8; DB 3; Length 2128;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1812; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY	1	gatgtgctctccaccctccctctgctgctcctcaccctgctccatctgtagcg	60
DB	294	gatgtgctctccaccctccctccctgctgctcctcaccctgctccatctgtagcg	353
QY	61	atgttcaatgaagaagactctgcacgtgcatctatctgtgctcctggtctatctg	120
DB	354	atgttcaatgaagaagactctgcacgtgcatctatctgtgctcctggtctatctg	413
QY	121	ctgtgctcactgtgctgctgaactctcttcaggttggaagtgtgactgtacactg	180
DB	414	ctgtgctcactgtgctgctgaactctcttcaggttggaagtgtgactgtacactg	473
QY	181	ggtctgagctcagaagaaatctcaagcagctgtaggaatctgtataatttctg	240
DB	474	ggtctgagctcagaagaaatctcaagcagctgtaggaatctgtataatttctg	533
QY	241	aaacttcagaagaagagctctacactgttcacaggtgcacccagggagacaaagagca	300
DB	534	aaacttcagaagaagagctctacactgttcacaggtgcacccagggagacaaagagca	593
QY	301	gtcttcagctatctgaataacctgaggtcagaagaagcgagagccttcacagac	360

Db 594 gtgtctcaagcttcttgaataacctggaaggtcaagaagaagcgaagccttcaagac 653
QY 361 acccaataacctctccctaccagaagactgtgagcaacttcaaggctgaagaagttcata 420
Db 654 acccaataacctctccctaccagaagactgtgagcaacttcaaggctgaagaagttcata 713
QY 421 cagttcccaactgagcaagaagaggtgaggttccctattgcatactcatatggtatcat 480
Db 714 cagttcccaactgagcaagaagaggtgaggttccctattgcatactcatatggtatcat 773
QY 481 gagaagattgaaaacttgaagaagctactgcgagctgtgtatgctccctgaagacataac 540
Db 774 gagaagattgaaaacttgaagaagctactgcgagctgtgtatgctccctgaagacataac 833
QY 541 tggatcatatgtgatatgagaagtcctccagaacttcaaaaggcgtcaagaacattatc 600
Db 834 tggatcatatgtgatatgagaagtcctccagaacttcaaaaggcgtcaagaacattatc 893
QY 601 tcttgcttcccaaatgtctcatatagccaagctggttcggttggtttaatgctccctg 660
Db 894 tcttgcttcccaaatgtctcatatagccaagctggttcggttggtttaatgctccctg 953
QY 661 tccagaggtgagaactgacactccaactgcatggaagactgtgtccagagctgaagtcgtg 720
Db 954 tccagaggtgagaactgacactccaactgcatggaagactgtgtccagagctgaagtcgtg 1013
QY 721 aactacttccctgatacatgtggagcgaacttccctataaaggacaaatgagagatgctc 780
Db 1014 aactacttccctgatacatgtggagcgaacttccctataaaggacaaatgagagatgctc 1073
QY 781 caggtctcacaagatgttgaatggagggaatagcatggaatcagaagtaacctcctaagac 840
Db 1074 caggtctcacaagatgttgaatggagggaatagcatggaatcagaagtaacctcctaagac 1133
QY 841 aagaagaacccgcgggaataatcactttagtgatgtagagacacactaacctcaaac 900
Db 1134 aagaagaacccgcgggaataatcactttagtgatgtagagacacactaacctcaaac 1193
QY 901 aagaagaagagatcctccctctataattactatgttttcaagggaatgcatcatgtg 960
Db 1194 aagaagaagagatcctccctctataattactatgttttcaagggaatgcatcatgtg 1253
QY 961 gcttcccgagattcgttccaacatgttttgaagaacacttaacccaacactgattgaa 1020
Db 1254 gcttcccgagattcgttccaacatgttttgaagaacacttaacccaacactgattgaa 1313
QY 1021 tgggttaaaagacattatagccagatgaacactcctgggccaaccttcagcgtgcag 1080
Db 1314 tgggttaaaagacattatagccagatgaacactcctgggccaaccttcagcgtgcag 1373
QY 1081 tggatgctgtgctctgttcccaacaccccaagtaacacatcctcagacatgacttcatc 1140
Db 1374 tggatgctgtgctctgttcccaacaccccaagtaacacatcctcagacatgacttcatc 1433
QY 1141 gccaggtcgtgtaagtggcagggtcatatgagagacatcgtataaggtgtccttatgct 1200
Db 1434 gccaggtcgtgtaagtggcagggtcatatgagagacatcgtataaggtgtccttatgct 1493
QY 1201 cctgtcttgaatccacccgggagctatctggtttatgggctgggagctggaattgg 1260
Db 1494 cctgtcttgaatccacccgggagctatctggtttatgggctgggagctggaattgg 1553
QY 1261 atgtctcaaaaacacatcactgttgcgcaacaagtttgaaccaaggtaagtataatgct 1320
Db 1554 atgtctcaaaaacacatcactgttgcgcaacaagtttgaaccaaggtaagtataatgct 1613
QY 1321 ctccagatgcttagaagaataactacgttataaggccaactatgaggactgaacttgagac 1380
Db 1614 ctccagatgcttagaagaataactacgttataaggccaactatgaggactgaacttgagac 1673
QY 1381 acactatgagagcgttgcctacactgttggggaagaagatgacaaaatgttcgaacttg 1440
Db 1674 acactatgagagcgttgcctacactgttggggaagaagatgacaaaatgttcgaacttg 1733

QY 1441 ctgggaagctgttgggtggagaccaggccttgcnaattcgttggcatcctttaagataaga 1500
Db 1734 ctgggaagctgttgggtggagaccaggccttgcnaattcgttggcatcctttaagataaga 1793
QY 1501 gggctgtatataatgttgggtgaagttagatcctttgcttgcnaaatgtcgtcgtgggtga 1560
Db 1794 gggctgtatataatgttgggtgaagttagatcctttgcttgcnaaatgtcgtcgtgggtga 1853
QY 1561 atgtctgttctctcaacccctaaccttagttagttccctcacttaacttctcaactgaat 1620
Db 1854 atgtctgttctctcaacccctaaccttagttagttccctcacttaacttctcaactgaat 1913
QY 1621 gagaatgagaacactgctgtgaataggagagtgaaagggaatagatgttgaagacattgat 1680
Db 1914 gagaatgagaacactgctgtgaataggagagtgaaagggaatagatgttgaagacattgat 1973
QY 1681 ttcaagttgaatgcctgctgtgtaagcttccatcctgttgaagctgcgcttccataatc 1740
Db 1974 ttcaagttgaatgcctgctgtgtaagcttccatcctgttgaagctgcgcttccataatc 2033
QY 1741 caggtttgtagcgttggaggagaacttggatggaagaagacacttccctctgtactgtt 1800
Db 2034 caggtttgtagcgttggaggagaacttggatggaagaagacacttccctctgtactgtt 2091
QY 1801 aacttaaaataaata 1816
Db 2092 aacttaaaataaata 2107

RESULT 2
US-07-955-041-3
Sequence 3, Application US/07955041
Patent No. 5360733

GENERAL INFORMATION:

APPLICANT: FUKUDA, MINORU
APPLICANT: BIERHUIZEN, MARTI FA
TITLE OF INVENTION: A NOVEL BETAL-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUCOSTALIN AND A METHOD FOR CLONING PROTEINS HAVING
TITLE OF INVENTION: ENZYMATIC ACTIVITY

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700

CITY: SAN DIEGO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/955,041

FILING DATE: 19921001

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATHRYN

REGISTRATION NUMBER: 31,815

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-535-9001

TELEFAX: 619-535-8949

INFORMATION FOR SEQ. ID NO. 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2105 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: both

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

```

: NAME/KEY: CDS
: LOCATION: 220..1504
: FEATURE:
: NAME/KEY: polyA_signal
: LOCATION: 1913..1918
: FEATURE:
: NAME/KEY: misc_signal
: LOCATION: 248..314
: OTHER INFORMATION: /standard_name=
: OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-07-955-041-3

```

```

Query Match          19.6%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 5.9e-111;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

```

Oy 337 aagaagcgaagccttcacagacccaccctccctccacagagactgtgagac 396
Db 466 AAAAAAGCCCTCGGTGACACCTGACGACTATATAAACATGACAGAGCTGTTCTT 525
Oy 397 ttcaagctgaagaagtcatacagtcctccactgagcaagaagtgagttccct 456
Db 526 TTCAATCAAGAGACGCAATATATTTGTAGAACCCCTTACTAAGAAGAGCGAGTTTCA 585
Oy 457 attcatactatgtgtatcatgagaagaattgaaccttgaaagactcagagact 516
Db 586 ATACGATATCTTATAGTGTTCATCACAAGATGAAATGCTTGACAGGCTGCTAGGGCC 645
Oy 517 gtgtatgccctcagaacatactgtgtcattgtgataggaagtcgccgaaccttc 576
Db 646 ATCTATATGCTCAGAAATTTCTATTTGCTTATGTCATGACACAAATCCGAGGATCTCTAT 705
Oy 577 aagaagcggcgaagaagcaattatcttctcccaaatgtctcatagccagtaagctg 636
Db 706 TTACGCTGAGTGAATGGGATGCTTCTCTGTTTATGATATGCTTTGTGTGGCAGCCGATTG 765
Oy 637 gtctgggtgttatttgcctcctgtgtccaggtgtcaagctgacccctcaactgtatgaaagac 696
Db 766 GAGAGTGTGTTTATGATCGTGGAGCGGGTTACAGGCTGACCTCACTGATGAAGAT 825
Oy 697 ttgtctcagagctcagtgccgtgtgaataacttccctgaatatactgtggaagcagacttccct 756
Db 826 CTCTATGCAATGAGTGCACAACTGGAAGTACTTGATTAATCTTTGTGGTATGATTTTCC 885
Oy 757 ataaagagcaatgcaagatgtgtccagcttcaagatgttgatggagaataagactg 816
Db 886 ATTAACCAACCTAGAAATTTGTCAAGAGCTCAAGTTGTTAATGGAGAAACCAACTG 945
Oy 817 gactcagaggtactcctcctaaagacaaacccgtgtgaaataacttgaagtgatg 876
Db 946 GAAAGGAGAGGATGCGCATCCCATTAAGAAAGAGAGTGAAGAGCGGTATAGAGTCTT 1005
Oy 877 agagaacattacactaaacaaagaagaagatcctccctcctataacttaactatg 936
Db 1006 AATGGAAGCTGACAAACACAGGAGCTGTCAAAATGCTTCCCTCCACTGGAACACCTCTC 1065
Oy 937 tttaagggatgtgtatctgtgtgtcctccagagatttcgtccaaatglttgaagac 996
Db 1066 TTTCTGGCACTGCTTACTTGTGTGTCAGTAGGAGTAGTGGGTATGACTACTACAGAAAT 1125
Oy 997 cctaactcccaacactgtatgaaatggtlaaaagacacttlaagccagatgaacactc 1056
Db 1126 GAAAAAATTCAAAAGTTGATGAGAGTGGGACACAAAGACATATACAGCCCTGATGAGATCTC 1185
Oy 1057 tgggacacccctcagcgtgtcaagtgatgtgcctgtgtccttcccaacaccccaaglac 1116
Db 1186 TGGGCACTGATCAAAAGATTCCTGAAGTCCCGGCTCAGTCCCGCAGCAGCATTAAGTAT 1245
Oy 1117 gacatctagacagactctatgtccagcgtgtgtcaagtgatgagcaggtgtcaatgagaagac 1176
Db 1246 GATCTATCTAGCATGCAAGAGGTGCGAGGTTTGTCAAGTGGCAGTACTTGTGAGGGTAT 1305

```

```

Oy 1177 atcgataaggtgtcctcctatgtcctccctgtctgtgaatccacacagcggtatctgcgt 1236
Db 1306 GTTCCAGAGGGTGTCCCTTACCCGCGCATGAGTGCATGTGTCCGCTCAGTGTGAT 1365
Oy 1237 tatgggctggggactgtgaatgtgacttcaaaacacactgtgtggcaacaagttc 1296
Db 1366 TTCGAGCTGTGACTTGTAACTGATGCTGGCAGCAACACCACTTGTGTGCCAATTAAGTTT 1425
Oy 1297 gaccacaaggtagatgaataagctcttcagtgcttagaagaatactcgttatagac 1355
Db 1426 GACGTGATGTTGACCTCTTGCATCCAGTGTGTGATGACATTTGAGACACCAAGC 1484

```

RESULT 3

US-08-227-455-3

```

: Sequence 3, Application us/08227455
: Patent No. 5624832
: GENERAL INFORMATION:
: APPLICANT: FUKUDA, MINORU
: APPLICANT: BIERHUIZEN, MARTI FA
: TITLE OF INVENTION: A NOVEL BETA1-6
: TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
: TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: CAMPBELL AND FLORES
: STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
: CITY: SAN DIEGO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/227,455
: FILING DATE: 14-APR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: CAMPBELL, CATHERYN
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9957
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-535-9001
: TELEFAX: 619-535-8949
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2105 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 220..1504
: FEATURE:
: NAME/KEY: polyA_signal
: LOCATION: 1913..1918
: NAME/KEY: misc_signal
: LOCATION: 248..314
: OTHER INFORMATION: /standard_name=
: OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-227-455-3

```

```

Query Match          19.6%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 5.9e-111;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```



```

Db 646 ATCTATATGCTCAGAAATTTCTATTGCTTCATGTGACACAAAATCCGAGATTCTTAT
Oy 577 aaagagcggtcaagaacatlaattcttgcctcccaaatgctcatalagcaagtaagctg 636
    || || || || || || || || || || || || || || || || || || || || ||
Db 706 TTAGCTGCAGTATGATGGCATCGCTTCCTGTTTACTAATGTCTTTGTGGCCAGCGGATTG 765
Oy 637 gtccgggtgttattatgctcctcgtgctcaggtgcaagctgcaactcaactgcaatgaaagc 696
    || || || || || || || || || || || || || || || || || || || || ||
Db 766 GAGAGTGTGTTTATGATCGTGAGCCGGGTTGAGGCTGACCTCAACTGATGAAGAT 825
Oy 697 ttgtcccaagagctcagtgccggtggaataactcctgtaactcgtggaagcaactcct 756
    || || || || || || || || || || || || || || || || || || || || ||
Db 826 CTCTATGCAATGATGTCACAACTGGAAGTATGATTAATCTTGTGATGATGATTTTCC 885
Oy 757 ataagaagaatgcagagatggtccagagctcgaagatgltgaatggagagatgagcatg 816
    || || || || || || || || || || || || || || || || || || || || ||
Db 886 ATTAACCAACCTAGAAATTTGTCAAGAACTCAAGTTTATATGGAGAAAACAACCTG 945
Oy 817 gagtcagaggtacccctcctaagcacaaagaacccgctggaatatacactttgaggtagtg 876
    || || || || || || || || || || || || || || || || || || || || ||
Db 946 GAAACGAGAGAGATGTCATCCATTAAGAAAGAAAGTGAAGACAGCGGTATGAGTGTCT 1005
Oy 877 agagacacattacactcaaccacaagaagaagatcctcccttataactatg 936
    || || || || || || || || || || || || || || || || || || || || ||
Db 1006 AATGGAAGCTGCACAAACACAGGACTGTCAAAATGCTTCCACCTGCAACACCTCTTC 1065
Oy 937 ttacagaggaatgctgacatggtgctccagagatllcccaacatgttttgaagaac 996
    || || || || || || || || || || || || || || || || || || || || ||
Db 1066 TTTTCTGCGACGTGCTACTTCTGTGTCAGTATGAGGATATGTGGGTATGTACTACAGAA 1125
Oy 997 cctaaaccacaacactgaltgaatggglaaagaacacttataagccagatgaacaactc 1056
    || || || || || || || || || || || || || || || || || || || || ||
Db 1126 GAAAAAATCCAAAGTTGATGAGTGGGCAACAAGACATACAGCCCTGATAGATATCTC 1185
Oy 1057 tgggccaaccttcagcgtgacggtgagatgectgctgtgtcccaaccacccaagtaac 1116
    || || || || || || || || || || || || || || || || || || || || ||
Db 1186 TGGGCGACCATTCACAAAGATTCCTGAAGTCCCGGCTCATCTCCGCGACCATTAAGTAT 1245
Oy 1117 gacatcagacatcactcactatgcccagcgtgacagtgacaggtgacatgaaggagac 1176
    || || || || || || || || || || || || || || || || || || || || ||
Db 1246 GATCTATCTGACATCAAGCACTTCCAGGTTTGTCAAGTGCATCTTATGAGGATGAT 1305
Oy 1177 atcgaataaggtgctcctatgctcctgctcgtcgtgacacacagcggtcactcgtcgt 1236
    || || || || || || || || || || || || || || || || || || || || ||
Db 1306 GTTTCAGAGGTGCTCCCTACCCGCGCATGAGTGCATGTCGCTCACTGTGCAAT 1365
Oy 1237 tatggagctggagacttgaaatgagatgcttcaaaacacacactcgttggccaagaattt 1296
    || || || || || || || || || || || || || || || || || || || || ||
Db 1366 TTCCGAGCTGTGATCACTGAACGTGATGCTGCACAAACACACTTGTGCTCAATTAAGTTT 1425
Oy 1297 gaccacaaggtagatgaataatgctcctcagtgcttagaagaatcactacgttataagc 1355
    || || || || || || || || || || || || || || || || || || || || ||
Db 1426 GAGGTGATGTTGACCTCTTGGCATCCAGTGTGTTGATGAGACATTTGAGACAAAAGC 1484

```

US-08-487-069-3

```

; Sequence 3, Application US/08487069
; Patent No. 5684134
;
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL, BETA1-6
; TITLE OF INVENTION: N-ACETYLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMACTIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA

```

```

; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,069
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ. ID NO.: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 220..1504
; FEATURE:
; NAME/KEY: POLYA_signal
; LOCATION: 1913..1918
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 248..314
; OTHER INFORMATION:
; OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
; US-08-487-069-3

```

Query Match 19.6%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 5.9e-11;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

Oy 337 aaagaagcgagagccttcaagaacacacactcctcctcacaagaagactgtaagc 396
    || || || || || || || || || || || || || || || || || || || || ||
Db 466 AAAAAAGCGCCCTCGGTGGACACCTGACGACTATTAATAACATGACCAAGTACGTTCTTCT 525
Oy 397 tccaagctgaaagaagtacatcacagttcccactgagaagaagaaggtggaattccct 456
    || || || || || || || || || || || || || || || || || || || || ||
Db 526 TTCAATCAGACAGACGCAAAATATTTGTAGAACCCCTTAGTAAGAGAGAGCGGATTTCCA 585
Oy 457 atgcatactcatalgtgaltcatgagaagatlgaaaacttgaagaagctactcgtgagct 516
    || || || || || || || || || || || || || || || || || || || || ||
Db 586 ATAGCATATTTATATAGTGGTTCATCACAAGATTGAATGCTGACAGGCTGCTGAGAGGC 645
Oy 517 gctatgccctcagaacatataactgltcactggtgagatgagaagtcccaaaaacttcc 576
    || || || || || || || || || || || || || || || || || || || || ||
Db 646 ATCTATATGCCCTCAGAAATTTCTATTGCGTTCACTGTGGACACAAAATCCGAGATTCTTAT 705
Oy 577 aaagagcggtcaagaacatlaattcttgcctcccaaatgctcatalagccagtaagctg 636
    || || || || || || || || || || || || || || || || || || || || ||
Db 706 TTAGCTGCAGTATGATGGCATCGCTTCCTGTTTAACTAATGTCTTTGTGGCCAGCGGATTG 765
Oy 637 gtccgggtgttattatgctcctcgtgctcaggtgcaagctgcaactcaactgcaatgaaagc 696
    || || || || || || || || || || || || || || || || || || || || ||
Db 766 GAGAGTGTGTTTATGATCGTGAGCCGGGTTGAGGCTGACCTCAACTGATGAAGAT 825
Oy 697 ttgtcccaagagctcagtgccggtggaataactcctcgaataactcgtggaagcaagcttcc 756
    || || || || || || || || || || || || || || || || || || || || ||
Db 826 CTCTATGCAATGATGTCACAACTGGAAGTATGATTAATCTTGTGATGATGATTTTCC 885

```


Db 1245 GATCATATGACATGACAGAGCTTCCAGCTTTGTCAAGTGCGACTCTTTCAGGGTGAT 1304
Qy 1177 atcgaagaaggtgctcctctatgctcctcgtctcgtatccaccagcgggctatcgtcgtt 1236
Db 1305 GTTTCAGAGGGGTGCTCCTTACCCGCCCTGCAGATGAGTGCATGTCGCTCATGTGTCATT 1364
Qy 1237 tatgggctggggagcttgaattgagtgcttcaaaacacatcactgcttggcacaagttt 1296
Db 1365 TTCGAGAGCTGAGTGAAGTGAAGTGCAGCAACACCACTGTTGGCAATTAACCTT 1424
Qy 1297 gaccacaaggtagatgataatgctctcaagtgcttagaagaatacctactgataagc 1355
Db 1425 GAGTGCATGTTGACCTCTTGGCCATCCAGTGTTCATGACCAATTGGAGACACAAAGC 1483

RESULT 7
US-08-118-906-13
Sequence 13, Application US/08118906
Patent No. 5484590
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,906
FILING DATE: 09-SEP-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1807 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 255..1454
US-08-118-906-13

Query Match 9.6%; Score 180.4; DB 1; Length 1807;
Best Local Similarity 54.3%; Pred. No. 7.3e-49;
Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

Qy 388 tgtgagcacttcaaggtcgtgaagaggttccacgtcagagaagaagagtg 447
Db 468 TGCAAGAGATCTTGACCCAGAGCACTACACAGCCCTTTATCAAGAAAGAGCT 527
Qy 448 gagttccctatgtcactctatggtgattcaatgagaagattgaaacttgaaagcta 507
Db 528 GACTTCCCTTGCGATATATATGATGTCATCATCTTTCAGACCTTTCAGAGGCTC 587

Qy 508 ctgcagcgtgtatgccccctcagaacatactgtgtccatgtgatgagatgccca 567
Db 588 TTCAGAGGCTATTACATGCCCCAAATATCTACTGTGTCATGATGATTAAGAAACAA 647
Qy 568 gaaacttcaagaagcggtgcaagaactatcttctgcttcccaaatgtcttcatagcc 627
Db 648 ACTGAATTTAAGATGCGGTAGACCACTATTAACTCTTCCCAACCGCTTTCTGCGCT 707
Qy 628 agtaagctggttcggtgtgtatgtctctcctcgttcaggggtgcaagctgactcaatc 687
Db 708 TCCAAGATGGAACCCGTTGTCATGAGAGGATCTCCAGGCTCCAGGCTGACCTGAAC 767
Qy 688 atggaagactgtctcagagctcaggtccggtggaatcttcctgaatacagtgagg 747
Db 768 ATCAGAGATCTTTCGCTTCAGAGTCTCATGGAATGAGTATATACACCTGTGGCCA 827
Qy 748 gacttccctataaagaacatgcaagatggtccaggtctcaagatgttgaatggaag 807
Db 828 GACTTCCCGCTGAACCAAGCAAGAAATAGTTCAGTATCTGAAGATTTAAAGGTAA 887
Qy 808 aatgcatgaggtcagaggtactctcctaagcacaagaaccgctggaatat----- 861
Db 888 AATATCACCCCGAGGGGTGCTGCCAGCTCATGCAATTTGACGAGCACTAAATATGTCAC 947
Qy 862 -----cacttgaggtgtagagagacacattacacactcaacaagaagaagatcct 915
Db 948 CAAGAGCACCTGGCAAAAGCTTTCTATGTGATGAACACACCGTTGAACCGCT 1007
Qy 916 ccccttataatctaactatgtttacaggaatggtcatgtgtgtctccagatctc 975
Db 1008 CCCCCCAATCTCATCAATTTACTTTGGCTGCTGCTATGTGGCTATCAAGAGAGTT 1067
Qy 976 gtccaatgttttgaagaaccttaalcccaactgatgtgataaagaacct 1035
Db 1068 GCAACTTTGTTCTCATGACCCACCGGCTGTGATTTGCTCCAGTGCGCAAGAGACCT 1127
Qy 1036 tatagccagatgaacacctctggccacctcagcgtgacagtgtagctgctc 1095
Db 1128 TTCAGTCCGTGATGACATTTCTGGGTGACACTCAATAGATTCCAGGTCTTCTGCTCT 1187
Qy 1096 gtcccaaa 1103
Db 1188 ATGCCAAA 1195

RESULT 8
US-08-486-196-13
Sequence 13, Application US/08486196
Patent No. 5731420
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,196
FILING DATE:
CLASSIFICATION: 424

```

? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/118,906
? FILING DATE: 09-SEP-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Campbell, Cathryn A.
? REGISTRATION NUMBER: 31,815
? REFERENCE/DOCKET NUMBER: P-LJ 9526
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619) 535-9001
? TELEFAX: (619) 535-8949
? INFORMATION FOR SEQ ID NO: 13:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1807 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 255..1454
? US-08-486-196-13

```

Query Match	Similarity	9.6%	Score 180.4	DB 1	Length 1807
Best Local	Similarity	54.3%	Pred. No. 7,3e+49		
Matches	395	Conservative	0	Mismatches	321
				Indels	12
				Gaps	1
QY	388	tgtgagcacttcaagagcttgaagaagatcatalacagttcccaactgagcaagaagagtg	447		
Db	468	TGCAAGGAATACTTATACCCAGACCACTACATCACAGCCCCCTTTATCTAAGGAAGAAGCT	527		
QY	448	gagttccattgagcactactatgtgtatctatgtgaagatgtgaacttgaagagcta	507		
Db	528	GACTTTCCCTTGGCATATTAATGATCCATCATCACTTTGACACCTTTGGAAAGGCTC	587		
QY	508	ctgcgagctgtgtatgccccctcagaacataatctgttccatgtgtatagaagttccca	567		
Db	588	TTCAAGGGCTATTACATGCCCCCAAAATATCTACTGTGTCAATGTGATGATAAAAGCAACA	647		
QY	568	gaacttccaaagagcggttcaagaacatatttctgtctcccaaatgtcttcatagcc	627		
Db	648	ACTGAATTTTAAAGATGAGCGGTAGAGCAACTATTTAAGCTGTCCCAACCTTTTCGGCT	707		
QY	628	agtaagctgtgttcggtgtgtatgttccttccttcgttcgaaggtgcaagctgaccactgc	687		
Db	708	TCCAAGATGGAACCCGTTGTCTATGTGAGGATCTCCAGGCTCCAGGCTTACCTGAACTGC	767		
QY	688	atggaagactgtctccagagctcaagtcagttgcgtgtgaataatcttctaatacatgttggag	747		
Db	768	ATCAGAAGATCTTTTCGCTTCGAGGTCATCATGGAAGTACGTTATCAACACCTGTGGGCAA	827		
QY	748	gacttctctataaagagcaatgtcagaagatgtgtcccaagcctctaagaatgttgtaatggag	807		
Db	828	GACTTCCCTCCCTTAACCAACGAAGAATAGTGTACGTATCTGAAGAAGATTTAAAGGTAA	887		
QY	808	aatagcatgtgagatcagagttacctctctaagcacaagaacccgctgtgaatat-----	861		
Db	888	AATATACCCCAAGGGGTGCTGCCCCAGCTATGCAATTGGACGCACTAAATATGTCCAC	947		
QY	862	-----cactttaggtatgtgagagacacatatacactaaccaacaagaagaatcct	915		
Db	948	CAMGAGCACTGTGGCAAGAGACTTTCTATGTGATTAAGAACAAACAGCGCTTGAACACGGCT	1007		
QY	916	cccccttaatttactactatgtttcaaggaatgtgtacatgtgtgtcccgagattc	975		
Db	1008	CCCCCCCCAATACCTACAAATTACTTTTGGCTGTGCTCATGTGAGCTTCAAGAAGAGTTT	1067		
QY	976	gtccaacaatgttttgaagaacacttaattccacaacactgtatgtgtttaaagaagcact	1035		
Db	1068	GCAACTTTTGTCTCATATGACCCACAGGGGCTTTGATTGTCTCCAGTGGTCCAAAGGACACT	1127		
QY	1036	tatagcccaagatgaacacctctgggccaaccttcagcggtgcaacggtgtatgacctgacct	1095		

Db 1128 TTCACTCTCATGAGCATTTCTGGGTGACACCTAAATAGAGATTCAGAGTGTCTCTGGCTCT 1187

QY 1096 gttcccaa 1103
| | | |

Db 1188 ATGCCCAA 1195

```

1 RESULT 9
2 US-08-488-135-13
3 Sequence 13, Application US/08488135
4 Patent No. 5766910
5
6 GENERAL INFORMATION:
7 APPLICANT: Fukuda, Minoru
8 APPLICANT: Bierhuizen, Marti F.A.
9 TITLE OF INVENTION: Expression of the Developmental I
10 TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
11 TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
12 NUMBER OF SEQUENCES: 14
13
14 CORRESPONDENCE ADDRESSES:
15 ADDRESSEE: Campbell and Flores
16 STREET: 4370 La Jolla Village Drive, Suite 700
17 CITY: San Diego
18 STATE: California
19 COUNTRY: USA
20
21 ZIP: 92122
22
23 COMPUTER READABLE FORM:
24 MEDIUM TYPE: Floppy disk
25 COMPUTER: IBM PC compatible
26 OPERATING SYSTEM: PC-DOS/MS-DOS
27 SOFTWARE: PatentIn Release #1.0, Version #1.25
28
29 CURRENT APPLICATION DATA:
30 APPLICATION NUMBER: US/08/488,135
31 FILING DATE:
32 CLASSIFICATION: 424
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/118,906
35 FILING DATE: 09-SEP-1993
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Campbell, Cathryn A.
38 REGISTRATION NUMBER: 31,815
39 REFERENCE/DOCKET NUMBER: P-LJ 9526
40
41 TELECOMMUNICATION INFORMATION:
42 TELEPHONE: (619) 535-9001
43 TELEFAX: (619) 535-8949
44
45 INFORMATION FOR SEQ ID NO: 13:
46 SEQUENCE CHARACTERISTICS:
47 LENGTH: 1807 base pairs
48 TYPE: nucleic acid
49 STRANDEDNESS: double
50 TOPOLOGY: linear
51 MOLECULE TYPE: cDNA
52 FEATURE:
53 NAME/KEY: CDS
54 LOCATION: 255..1454
55
56 US-08-488-135-13

```

	Query Match	9.6%	Score 180.4	DB 1	Length 1807
	Best Local Similarity	54.3%	Pred. No. 7.3e-49		
	Matches 395	Conservative 0	Mismatches 321	Indels 12	Gaps 1
QY	388	tgtagcaactcaagagcctgtaagaagagtcatactacagttcccaactgagcaagaagagtg	447		
Db	468	TGCAAGGAATACCTTACCACAGCCATCAACAGGCCCTTTTCTAAGAGAAGACT	527		
QY	448	gagttcccatatgcaactatagtgatgcatgaggaagatgaaactttgaaagcta	507		
Db	528	GACCTTCCCTGGCATATATATATGCTCATCCATCATCTTTGACACCTTTGCCAAGGCTC	587		
QY	508	ctgagagctgtagtaccgcccctcagaacatactatgctccatgtagtagagaagctccca	567		
Db	588	TTCAAGGCGCTATTTAATGATCCCCCAAAATATCTACTGTTCATGTCAGTGAAGAAAAGCAACA	647		

```

QY 568 gaacttcaagaagcggtcaagaacattatcttctccccaatgctcagacc 627
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 ACTGAATTTAAGATGCGGTAGAGCACTATTAGCTCTTCCCAAGCTTTTGGCT 707
QY 628 agtaagcttggtcggtgttatacgtctcgtcgtcgtcgtcgtcgtcgtc 687
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 TCCAAGATGGAACCGGTGTCTATGAGAGGATCTCCAGGCTCCAGCTGAACTGC 767
QY 688 atggaagactgctcccaagctcagtcggtcggtggaatctccctgaatactggtgagc 747
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 768 ATCAGAGATCTTTCTGCTCTGAGGTCTCATAGAGTATGCACTGATGAGGCA 827
QY 748 gacttccataaagaagaatgcaagaatggtccaggtcctcaagaatgttgaatggaag 807
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 828 GACTTCCCGCTGAAACACACAGGAATAGTTCAGTATCTGAAAGATTTAAAGCTAAA 887
QY 808 aatagatgagtcaggaagttacctcctcaagaacaaagaccgctggaatat----- 861
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 888 AATATCACCCCGAGGGGTGCTGCCCCAGCTCATGCAATTGAGCGACTAATATATGCCAC 947
QY 862 -----cacttgagtgtagagagacacattacacccaacaagaagaagactct 915
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 948 CAGAGCACTGGGCAAGAGCTTTCTATGATAGAACACAGCTTGAAACCGCT 1007
QY 916 cccctataatctaactatgtttacaaggaatggtacatgtgtgtcctccagatctc 975
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1008 CCCCCCATTAATCTACAAATTACTTTGGCTCTGTGCGCTATGAGTCTTCAAGAGATT 1067
QY 976 gtccaacatgtttgaagaaccctaataatccaacaacatgattgaatggttaagaacac 1035
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1068 GCCAACTTTGTCATGACATGACCCAGCGCTGTGATTTGCTCCAGTGTCCAAAGACACT 1127
QY 1036 tatagcccaagtagaacactcttggtccacactcagcgtgcagtgatgctcgtcct 1095
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1128 TTCAGTCTGTAGAGCAATTTCTGGGTGACACTCAATAGAGATTCCAGGTGTTCTGGCTCT 1187
QY 1096 gtcccaaa 1103
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1188 ATGCCAAA 1195

```

```

RESULT 10
US-08-474-065-13
: Sequence 13, Application US/08474065
: Patent No. 5830465
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Minoru
: APPLICANT: Biehuizen, Marti F.A.
: TITLE OF INVENTION: Expression of the Developmental I
: TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
: TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,065
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/118,906
: FILING DATE: 09-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.

```

```

: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9526
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1807 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 255..1454
: US-08-474-065-13

Query Match 9.6%; Score 180.4; DB 2; Length 1807;
Best Local Similarity 54.3%; Pred. No. 7.3e-49;
Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

QY 388 tgtgacacccaaggtcgaagaagtcacacagtcctccactgagcaagaaggtg 447
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 TCGAAGGATATCTTGACCCAGAGCCACTACATCACGCCCTTATCTAAGGAAGAGCT 527
QY 448 gattccctatgtctactctatcgtgtgtatcattgaaagatgaaacttgaaggtc 507
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 GACTTCCCTTGGCAATTAATATGATGTCATCCATCACTTTGACACCTTTCAAGGCTC 587
QY 508 ctgagctgtgtatgccccctcagacatatctgttccatgtgtatgagaagtcacca 567
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 TTCAGGGCTATTTACATGCCCCCAAAATATCTACTGTGTCATGATGATGAAAAACACA 647
QY 568 gaacttcaagaagcggtcaagaacattatcttctccccaatgctcagacc 627
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 ACTGAATTTAAGATGCGGTAGAGCACTATTAAGCTGCTTCCAAAGCTTTTCTGGCT 707
QY 628 agtaagcttggtcggtgttatacgtctcgtcgtcgtcgtcgtcgtcgtcgtc 687
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 TCCAAGATGGAACCGGTGTCTATGAGAGGATCTCCAGGCTCCAGGCTGAACTGC 767
QY 688 atggaagactgctcccaagctcagtcggtcggtggaatctccctgaatactggtgagc 747
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 768 ATCAGAGATCTTTCTGCTCTGAGGTCTCATAGAGTATGCACTGATGAGGCA 827
QY 748 gacttccataaagaagaatgcaagaatggtccaggtcctcaagaatgttgaatggaag 807
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 828 GACTTCCCGCTGAAACACACAGGAATAGTTCAGTATCTGAAAGATTTAAAGCTAAA 887
QY 808 aatagatgagtcaggaagttacctcctcaagaacaaagaccgctggaatat----- 861
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 888 AATATCACCCCGAGGGGTGCTGCCCCAGCTCATGCAATTGAGCGACTAATATATGCCAC 947
QY 862 -----cacttgagtgtagagagacacattacacccaacaagaagaagactct 915
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 948 CAGAGCACTGGGCAAGAGCTTTCTATGATAGAACACAGCTTGAAACCGCT 1007
QY 916 cccctataatctaactatgtttacaaggaatggtacatgtgtgtcctccagatctc 975
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1008 CCCCCCATTAATCTACAAATTACTTTGGCTCTGTGCGCTATGAGTCTTCAAGAGATT 1067
QY 976 gtccaacatgtttgaagaaccctaataatccaacaacatgattgaatggttaagaacac 1035
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1068 GCCAACTTTGTCATGACATGACCCAGGCTGTGATTTGCTCCATGAGGCAAGCACT 1127
QY 1036 tatagcccaagtagaacactcttggtccacactcagcgtgcagtgatgctcgtcct 1095
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1128 TTCAGTCTGTAGAGCAATTTCTGGGTGACACTCAATAGAGATTCCAGGTGTTCTGGCTCT 1187
QY 1096 gtcccaaa 1103
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1188 ATGCCAAA 1195

```


Db 133 CATGTGACACAAAATCCGAGATTCTATTAGCTGCAGTGTGGCATCCGCTTCCTGT 192
Qy 607 ttcccaaatgtcttcataagccagtaagctggttcgggtgtttatagtccttcgtccag 666
Db 193 TTTAATTAATGCTTTGTGGCCAGCGATTGGAGAGTGTGTTATGCAATCGTGGAGCCGG 252
Qy 667 gtgcagctgacccaactgcatggaagactgtctccagagctcagtcgcgtggaataac 726
Db 253 GTTCAGGCTGACCTCACTGATGATGAAGATCTCTATGCAATGAGTGCMAATGGAAGTAC 312
Qy 727 ttccgaatacatgtggtgagcagacttctctataaagagcaatgcagagatgttccagct 786
Db 313 TTGATTAATCTTTGTGTGATGATTTTCCCATTAACCACTGAAATTTGTCAAGAG 372
Qy 787 ctcaag 792
Db 373 CTCAG 378

RESULT 13

US-08-488-135-3
; Sequence 3, Application US/08488135
; Patent No. 5766910

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488.135
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA

NAME/KEY: CDS

LOCATION: 1..378

US-08-488-135-3

Query Match 8.6%; Score 162.8; DB 1; Length 378;
Best Local Similarity 65.3%; Pred. No. 1.5e-43;
Matches 239; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 427 ccactgagaagaagagtgagttccctatgcatactctatgtgtatcatgagaag 486
Db 13 CCCCTTAGTAAGAAAGAGCGGAGTTTCCATATAGCAATATTATAGTGTTCATGCAG 72
Qy 487 attgaaactttgaaagctactcagagctgtgtatgcccctcagaacatatactgtgc 546
Db 73 ATTGAATTCCTTGACAGGCTGCTGAGGGCCATCTATATCCCTCAGAAATTTATTCGGTT 132
Qy 547 catgtgagatgaagttcccccagaacttccaagagggcggtcaagaacttatctctgc 606
Db 133 CATGTGACACAAAATCCGAGATTCTATTAGCTGCAGTGTGGCATCCGCTTCCTGT 192
Qy 607 ttcccaaatgtcttcataagccagtaagctggttcgggtgtttatagtccttcgtccag 666
Db 193 TTTAATTAATGCTTTGTGGCCAGCGATTGGAGAGTGTGTTATGCAATCGTGGAGCCGG 252
Qy 667 gtgcagctgacccaactgcatggaagactgtctccagagctcagtcgcgtggaataac 726
Db 253 GTTCAGGCTGACCTCACTGATGATGAAGATCTCTATGCAATGAGTGCMAATGGAAGTAC 312
Qy 727 ttccgaatacatgtggtgagcagacttctctataaagagcaatgcagagatgttccagct 786
Db 313 TTGATTAATCTTTGTGTGATGATTTTCCCATTAACCACTGAAATTTGTCAAGAG 372
Qy 787 ctcaag 792
Db 373 CTCAG 378

RESULT 14

US-08-474-065-3
; Sequence 3, Application US/08474065
; Patent No. 5830465

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474.065
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-474-065-3

Query Match 8.6%; Score 162.8; DB 2; Length 378;
Best Local Similarity 65.3%; Pred. No. 1.5e-43;
Matches 239; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 427 ccactgagcaagaaggggagtgagttccctattgcatactctatggtgattcagagag 486
DB 13 CCCCTTGTAAAGAGAGCGCGATTTCCATACATATTTAAAGTGTTCAACACAG 72
QY 487 attgaacttgaagagctactgagctggtatgccccctcagaacatactggtc 546
DB 73 ATTGAATGCTTGACAGGCTCTGAGGCCATCTATATGCTCGAATTTCTATGCTT 132
QY 547 catgtgatgagaagtcgcccaaaacttcaagaagcggtcagaacataattcttc 606
DB 133 CATGTGACACAAATCCGAGATTCCTATTAGCTCGAGATGAGGATCGTTCCTGT 192
QY 607 ttcccaatgcttcataagccagtaagctggttcgggtgttatgctcctctgccaag 666
DB 193 TTAGTAAATGCTTTGTGGCCAGCCGATTTGAGAGTGTGTTAATGCTGTGAAGCCGG 252
QY 667 gtgcaagctgagctcaactgcatgagaactgctccagaagctcagtgccgtggaatc 726
DB 253 GTTACGCTGACCTCAACTGATGAGATCTATGCAATGAGTGCNAATGTGAAGTAC 312
QY 727 ttcttgataatggtgagagacttccctataagaagcaatgacagatggtccagct 786
DB 313 TTGTAATATCTTTGTGTATGATTTCCCATTTAAACCACTTAATTTGTGAGAG 372
QY 787 ctcaag 792
DB 373 CTCAAG 378

RESULT 15

US-08-118-906-1
Sequence 1, Application US/08118906
Patent No. 5484590
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,906
FILING DATE: 09-SEP-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-118-906-1

Query Match 7.1%; Score 134; DB 1; Length 378;
Best Local Similarity 60.4%; Pred. No. 4.5e-34;
Matches 221; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 426 ccaactgagcaagaagagtgagttccctattgcatactctatggtgattcagagag 485
DB 12 CCCTTTATCTTAAGAGAGAGCTGACTTCCCTTGCAATATTAATGTTGTCATCATATCA 71
QY 486 gattgaaacttgaagagctactgagctggtatgccccctcagaacataactgtgt 545
DB 72 CTTTGACACCTTTGCAAGGCTCTTCAGGCTATTATACATGCCCCAATATCTCTGTGT 131
QY 546 ccatgtgatgagaagtcgcccaaaacttcaagaagcggtcagaacataattcttc 605
DB 132 TCATGTGATGAAAGCAACAACTGAATTTAAGATGCGGTAGAGCAACTATTAAGCTG 191
QY 606 ctcccaaatgcttcataagcaagctggttcgggtgttatgctcctctggtccag 665
DB 192 CTTCCCAAGCCTTTGTGCTTCCAAAGATGAACCCGTCTTATGAGAGGATCTCCAG 251
QY 666 gtgcaagctgagctcaactgcatgagaactgctccagaagctcagtgccgtggaata 725
DB 252 GCTCAGGCTGACCTGAACCTCATCAGAGATCTTTCGCTTGAGAGTCTCATGGAAGTA 311
QY 726 ctctcgaatacatgtgagagacttccctataagaagcaatgacagatggtccagc 785
DB 312 CGTTATCAACACCTGTGGCAAGACTTCCCTGAAAAACCAAGAAATAGTTCACTA 371
QY 786 tctcaa 791
DB 372 TCTGAA 377

Search completed: September 26, 2002, 04:25:42
Job time: 9072 sec

THIS PAGE BLANK (U8PT0)

[illegible]

XX New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
PS Claim 3; Page 50-51; 66pp; English.
XX
CC The present sequence represents a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 438 AA;

Query Match 100.0%; Score 2360; DB 21; Length 438;
Best Local Similarity 100.0%; Pred. No. 2.3e-224;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWQKRRLCOLHYLMALGCYMLATVALKLSFRLCDSDHILGLESRESOSQYCRNIIYNFL 60
DB 1 mvgwkrllcqhhywalgymllatvalklsfrlkdshlglesresgqycrnllynfl 60
QY 61 KLPKRSINCSGVTRGDOEAVLOAILNINLEVKKKRPFDTHTYLSLRDCEHFAERKFI 120
DB 61 klprkrsincsgvtrgdgeavllqailnnlevkkkrpfldthylsltrdcehfkærkfi 120
QY 121 QPFLSKEEVEFPFAYSMVHEKTEINFERLLRAYAPONTICVAVDEKSPTEFEAVKATII 180
DB 121 qplskseevefpfaysmvhekteinferrllrayaponticvavdeksptefkæavkati 180
QY 181 SCPPNFIASKLVRVYVYASMSRVQADLNCMEDLLQSSVPKRYPLNCGTDFPIKSAEMV 240
DB 181 scfpnvfiasklvrvyvvyasmsrvqadlnmedllqssvpkryplntcgdfpiksaemv 240
QY 241 QALKMLNGRMSMEVPPKIKETRMKTHFRVYVDTLHLTKKKKDDPPYNTLTMTGNAYIV 300
DB 241 qalkmlngrmsmsepkkiketrmkthfrvyvdtlhltnkkkddppynltmtgnayiv 300
QY 301 ASRDPFOHVLKNPKRSQOLTEWVKDTYSPDEHLNATLORARMPGVSVPNHPKXDISDWTSTI 360
DB 301 asrdpfvohvlnkpnksqoltewvkdtyspdehlntatlorarmpgvsvpnhpkxdisdwtsti 360
QY 361 ARLVKKQGHGIDIKGAPAPCSGIHQORALCVYGAGDLMNMLONHLLANKFDPKVDNA 420
DB 361 arlvkkqghgdidkgapapcsghiqoralcyygagdlmnmlonhllankfdpkvddna 420
QY 421 LOCLEFLRYKATYGTFL 438
DB 421 lqcleelyrkaitygtel 438

DE Human C2/4Gnt protein.
XX
KW Human; C2/4Gnt; UDP-N-acetylglucosamine; O-glycan biosynthesis;
KW O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 10..27
FT /note="putative transmembrane domain"
XX
PN MO200034449-A2.
XX
PD 15-JUN-2000.
XX
PF 03-DEC-1999; 99WC-DK00677.
XX
PR 04-DEC-1998; 98DK-0001605.
XX
PA (CLAU/) CLAUSEN H.
XX
PI Clausen H, Schwiientek T;
XX
DR WPI; 2000-423407/36.
DR N-PSDB; AAA48623.
XX
PT New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
PT probe for the detection of specified glucosaminyltransferase from
PT other species and related organisms
XX
PS Claim 20; Fig 2; 47pp; English.
XX
CC The present sequence is human UDP-N-acetylglucosamine:
CC N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
CC (C2/4Gnt). C2/4Gnt is the third member of the family of O-glycan
CC beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
CC based O-glycans on to oligosaccharides, glycoproteins and
CC glycosphingolipids. C2/4Gnt can therefore be used in the production of
CC appropriately glycosylated glycoconjugates with particular enzymatic,
CC immunogenic, or other biological or physical properties. The nucleotide
CC sequence is useful as a probe for the detection of C2/4Gnt from other
CC species and related organisms and for the recombinant production of
CC C2/4Gnt polypeptide. The nucleotide sequence was identified by analysis
CC of EST database sequence information. Oligonucleotides derived from EST
CC clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
CC from a human foreskin genomic pl library by 5' RACE PCR. RT-PCR was
CC performed using Coloz05 human cell line mRNA in order to produce cDNA
CC for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
CC has been implicated in tumour progression and metastasis.
XX
SQ Sequence 438 AA;

Query Match 100.0%; Score 2360; DB 21; Length 438;
Best Local Similarity 100.0%; Pred. No. 2.3e-224;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWQKRRLCOLHYLMALGCYMLATVALKLSFRLCDSDHILGLESRESOSQYCRNIIYNFL 60
DB 1 mvgwkrllcqhhywalgymllatvalklsfrlkdshlglesresgqycrnllynfl 60
QY 61 KLPKRSINCSGVTRGDOEAVLOAILNINLEVKKKRPFDTHTYLSLRDCEHFAERKFI 120
DB 61 klprkrsincsgvtrgdgeavllqailnnlevkkkrpfldthylsltrdcehfkærkfi 120
QY 121 QPFLSKEEVEFPFAYSMVHEKTEINFERLLRAYAPONTICVAVDEKSPTEFEAVKATII 180
DB 121 qplskseevefpfaysmvhekteinferrllrayaponticvavdeksptefkæavkati 180
QY 181 SCPPNFIASKLVRVYVYASMSRVQADLNCMEDLLQSSVPKRYPLNCGTDFPIKSAEMV 240
DB 181 scfpnvfiasklvrvyvvyasmsrvqadlnmedllqssvpkryplntcgdfpiksaemv 240

QY 241 QALKMLNGRNSMSEVPPKHKETRMKYHFEVVRDTLHLTNKKKDDPPYNLTMFTGNAYIV 300
|
Db 241 galkmlngnrnsesevppkhketrwkyhfevvrdtlhltnkkkddppynlmtftgnayiv 300
QY 301 ASRDFVQVHLKMPKSOQLIEWKDTYSPDEHLMAFLQARMMPGSVPNHPKXDISDMSI 360
|
Db 301 asrdfvqvhvlnkpkssqqliewkdtyspdehlwafclqarwmpgsvpnhpkxdisdmsi 360
QY 361 ARLVWOGHEGIDKGAAPPCSGTHORALICYAGDLMWMLQNHLLANKRDPKVDNA 420
|
Db 361 arlvwoghegidkgapapcsghlqaralcvygsadlmwmlqnhllankrtdpkvddna 420
QY 421 LQCLEEYLRYKAIYGTSL 438
|
Db 421 lqcleeylrykaiygtel 438

RESULT 3
AAB30518
ID AAB30518 standard; Protein: 438 AA.
XX
AC AAB30518;
XX
DT 06-MAR-2001 (first entry)
XX
DE Amino acid sequence of beta-1-6-N-acetylglucosaminyltransferase.
XX
KW Human: beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M; Inflammation;
KW membrane protein; branched sialyl Lex; L-selectin; immune reaction;
KW Inflammation; tissue rejection; tumour metastasis.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Modified-site 289 /note- "potential N-glycosylation site"
XX
PN US6136580-A.
XX
PD 24-OCT-2000.
XX
PE 19-JAN-1999; 99US-0233506.
XX
PR 19-JAN-1999; 99US-0233506.
XX
PA (BURN-) BURNHAM INST.
XX
PI Fukuda M, Yeh J;
XX
DR WPI: 2001-040238/05.
XX
DR N-PSDB: AAC62134.
XX
PT New C2GNT-M polypeptides having core 2, core 4 and I branching
PT beta-1-6-N-acetylglucosaminyltransferase activities for preparing
PT reagents useful for diagnosing, preventing or treating inflammation or
PT tumour metastasis -
XX
PS Claim 1; Fig 4; 25pp; English.
XX
CC The present sequence represents a human
CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
CC and I branching activities. It is designated C2GNT-M. C2GNT-M is a
CC membrane protein that is predominantly expressed in colon, small
CC intestine, trachea, stomach and thyroid, as well as in certain cancer
CC cell lines. C2GNT-M polypeptides may be used to prepare molecules having
CC highly branched sialyl Lex and L-selectins, which may be subsequently
CC used to modulate immune reactions, e.g. inflammation and tissue
CC rejection, and to prevent or inhibit tumour metastasis.
XX
SQ Sequence 438 AA;

Query Match 100.0%; Score 2360; DB 22; Length 438;

Best Local Similarity 100.0%; Pred. No. 2.3e-224;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYQWRKLCQHLHMLAGCMMLATYALKSFRLKCDSDHLLGLESRSSQCYRNLINYL 60
|
Db 1 mwqwrklcqlhlhmlagcmmlatvalklsfrlkcdsdhlglesrsqcyernllynl 60
QY 61 KLPKRSINCSGVTGDDQAVLQALINLNEVKKRREPFDTHYLSLRDCEHFKAERKPI 120
|
Db 61 klprksinscsvtrgddqavvlqalnlnevvkkrrpfdthylslrdcehfkaerki 120
QY 121 QPPLSKEVEYFPDIAYSMVTHEKTIENFERLLRAVYAPQNIYCVHDEKSPTEKVAKAI 180
|
Db 121 qpllskeveyfpdiaysmvthehtienferllrayvapnqlycvhdekspetkvaakai 180
QY 181 SCFPNVFTASKLVRYVYASMSRVQADLNCMBEQLLOSSVWKKFELTWCIGDFFIKSAENV 240
|
Db 181 scfpnvftasklvrvyasmsrvqadlncmbdlqssvwpkyflntcgtldfriknaenv 240
QY 241 QALKMLNGRNSMSEVPPKHKETRMKYHFEVVRDTLHLTNKKKDDPPYNLTMFTGNAYIV 300
|
Db 241 galkmlngnrnsesevppkhketrwkyhfevvrdtlhltnkkkddppynlmtftgnayiv 300
QY 301 ASRDFVQVHLKMPKSOQLIEWKDTYSPDEHLMAFLQARMMPGSVPNHPKXDISDMSI 360
|
Db 301 asrdfvqvhvlnkpkssqqliewkdtyspdehlwafclqarwmpgsvpnhpkxdisdmsi 360
QY 361 ARLVWOGHEGIDKGAAPPCSGTHORALICYAGDLMWMLQNHLLANKRDPKVDNA 420
|
Db 361 arlvwoghegidkgapapcsghlqaralcvygsadlmwmlqnhllankrtdpkvddna 420
QY 421 LQCLEEYLRYKAIYGTSL 438
|
Db 421 lqcleeylrykaiygtel 438

RESULT 4
AAG75058
ID AAG75058 standard; Protein: 465 AA.
XX
AC AAG75058;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SPQ ID NO:5822.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 15.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PE 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
XX
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
XX
DR N-PSDB: AAH34463.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 7338-7340; 9803pp; English.
XX
SQ AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

50 Sequence 465 AA;

Query Match	100.0%;	Score 2360;	DB 22;	Length 465;
-------------	---------	-------------	--------	-------------

Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	MYWJRRCQLHYHMAAGCYMLLATYALKSTRKCDSPHLEBESRESQYCRNMLYNFL	60
Db	28	mqwvtrlqclhyhwalgcymllatvalklstrlcsdshlglesresqycrnllynfl	87
Oy	61	KLPAKRSINCSGVWGDEAVLADLAINLLEFKKKREPPTDNYSLFNDCEHFAEKREI	120
Db	88	klpaktrsincsgvtrtqgdgavlgallmlewkkrreptdchylstrlcehfaekrfi	147
Oy	121	QFPLSKEVEERPIATSWITHKEIENEFELRAVTAPONTYVHDEKSPETFKAVKAI	180
Db	148	qflpskeveerpiatamwihkeienferlraavyapqnlvcvhdekspefkavavaii	207
Oy	181	SCFPNWFASKLIVRVVYASMSRVADLNCMEDLLOSSVPKRYFLNTGCTDPRFISKNEMV	240
Db	208	scfpnwfasklrvrvvyaswsvrvgadlnccmedllqssvpykryflntgcfdprfkisnmemv	267
Oy	241	QALKMLNGRSMSESEVPRKHEKTRMKYHEEYVADTLHLTNKKKDRPYNLTMTGNAYIV	300
Db	268	qalkmlngrmsesevprkheketrwkyheevvadtllhlnkkkdrppynltmftgnayiv	327
Oy	301	ASRPDVGVHVLKNPSSQOLIEWVKOPVSPDEHLLMTLQARMPGSSVPRNRPYUJSDMTSI	360
Db	328	asrtdvqvhlknpsqglievwkddyspdenhlwllqatamprgsvprnryudjdsmtsl	387
Oy	361	ARLVKMOGHEDIDKGAFAAPCSGIHOARICIVUYGAGDINMMIQNHNLIDANKFDEKVDVDA	420
Db	388	arlvkmgghedidkgarparcsglhnqraicvuygagdlinnmlqnhhllankfdkrvddna	447
Oy	421	LQCEEFERYKATYGTGL 438	
Db	448	lqceeylrkyaiygtel 465	

RESULT	5
AA18996	
ID	AA18996 standard; Protein; 663 AA.
XX	
AC	AA18996;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	A partial core 2 beta-1,6-N-acetylglucosaminyltransferase.
XX	
KW	Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
KW	cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW	rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW	septic shock; adult respiratory distress syndrome; ARDS; cancer;

KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis.

OS Homo sapiens

PN CA2296936-A1.

PD 03-AUG-2000.

PF 03-FEB-2000; 2000CA-2296936.

PR 03-FEB-1999; 99US-0118674.

PA (GLYC-) GLYCODESIGN INC.

PI Korczak B, Lew A;

DR WPI; 2000-594746/57.

PT New nucleic acid molecules of core 2

PT compositions for treatment of disorders mediated by the enzyme

[illegible]

xx
xx
E

CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.

CC cancer, cardiovascular disorders and inflammatory disorders including
CC cancer, rheumatoid arthritis, inflammatory bowel disease,
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.

SQ Sequence 663 AA;

Query match	100.0%;	Score 2360;	DB 21;	Length 663;
-------------	---------	-------------	--------	-------------

Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y	1	mwkmrqlcohlwmalgcymllatvalvklstrlmcddshlgleseresosoqcrnllnyfl	60
Db	85	mwqmrictqlhylwalgcmllacvalklstrlcsdshlgleseresqscrnllnyfl	144
0y	61	klpaksrincsgvtrgddeavloaillnnlevrkkkrreepftdhrlysltrdcehfkraekrfi	120
Db	145	klpaksrincsgvtrgddeavlgallmlekkkrreftdhrlysltrdcehfkraekrfi	204
0y	121	qfplskseevefrplyaswihiekienferllirayapqnlcyvhdekspetfkxavkail	180
Db	205	qfplskseevefrplyaswihiekienferllirayapqnlcyvhdekspetfkxavkail	264
0y	181	scfprvfrlasklrvvyvaswrsrvadlnccmedllqssvpmkyflntcgtfdipklsnaemv	240
Db	265	scfprvfrlasklrvvyvaswrsrvadlnccmedllqssvpmkyflntcgtfdipklsnaemv	324
0y	241	qalkklrngrnsmesevpprkhetrmkxhfevayrdtlltlnkkkdpnylnltmfngnayiv	300
Db	325	qalkklrngrnsmesevpprkhetrmkxhfevayrdtlltlnkkkdpnylnltmfngnayiv	384
0y	301	asrdevoahlvknpkssooliemvkdjtsbdehlwattloqarhmmpgsvbnhrkydisdmfji	360

Db 365 asdfvghvlnkpsqqllewkdtyspdehlwatlqratwmpgsvnpbkydisdmtsi 444
 QY 361 ARLVKMOGHEGDIDKGAAPYAPCSGITHORALCYGAGDINMMLONHLLANKRDPVDDNA 420
 Db 445 arlvkwghegdidkgayapcsghgralcvygsqdinmlqnhllankfdpkvddna 504
 QY 421 LQCLEEYLRKAIYGTTEL 438
 Db 505 lqcleeylrkaiygtel 522

RESULT 6
 AAB54344
 ID AAB54344 standard; Protein: 465 AA.
 AC AAB54344;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:796.
 XX
 KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW nocrotic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.
 XX
 KM Homo sapiens.
 OS
 PN MO200055320-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05989.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-579444/54.
 XX
 DR N-PSDB; AAC99109.
 XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX
 XX
 PS Claim 11; Page 1237-1239; 1379pp; English.
 XX
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, nocrotic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent

CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 465 AA;
 Query Match 99.7%; Score 2354; DB 21; Length 465;
 Best Local Similarity 99.8%; Pred. No. 1e-223;
 Matches 437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVQWKRICOLHYIMALGCMYMLATVALKLSPLKCDSDHGLGESRESOSYCRNLYNPL 60
 Db 28 mvqwkrlcqlhyilwalgcmllatvalklsfklcdshlglstresgqcrnllnfl 87
 QY 61 KLPKRSINSGVTRGDPAVLOALINMLEVKKRREPETHYLSLTFDCEHFAERKEFI 120
 Db 88 klpkrsincsgvtctgdgavlgallmlevkkrreptdthylsttcdcehfkcerfll 147
 QY 121 QPPLSKEEVEPPIAYSMWIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEKIEK 180
 Db 148 qplskseevefpiaysmwvhekiekiekiekiekiekiekiekiekiekiekiekie 207
 QY 181 SCFPPVFIASKLVRYVYASWSRVQADLMCMEDLQSSVPWKYELNWTGDPPIKSNAEV 240
 Db 208 scfpnvfiasklvrvyaswsvrgadlnomedllqssvpwkyflntcgldpriknaev 267
 QY 241 QALKMLNGRNSMESEVPPKHKETPMKYHFEVVRDTLHLTNKKKDDPPYNLTFGTNAYIV 300
 Db 268 qalkmlngnrnsmevppkhketrkkyhfevvrtdlhltnkkkddppynlntfgnayiv 327
 QY 301 ASRDFVQHVLEKNPKSQQLIEWVKDTYSPDEHLMATLQRRMMPGSVPNHPKYDISDMTSI 360
 Db 328 asdfvghvlnkpsqqllewkdtyspdehlwatlqratwmpgsvnpbkydisdmtsi 387
 QY 361 ARLVKMOGHEGDIDKGAAPYAPCSGITHORALCYGAGDINMMLONHLLANKRDPVDDNA 420
 Db 388 arlvkwghegdidkgayapcsghgralcvygsqdinmlqnhllankfdpkvddna 447
 QY 421 LQCLEEYLRKAIYGTTEL 438
 Db 448 lqcleeylrkaiygtel 465

RESULT 7
 AAB18999
 ID AAB18999 standard; Peptide: 406 AA.
 XX
 AC AAB18999;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE A core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region.
 XX
 KW Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; AIDS; cancer;
 KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis.
 KW
 XX
 OS Homo sapiens.
 PN
 XX CA2296936-A1.
 PD
 XX 03-AUG-2000.
 PF
 XX 03-FEB-2000; 2000CA-2296936.
 PR
 XX 03-FEB-1999; 99US-0118674.
 PA (GLYC-) GLYCIDESEIGN INC.
 XX
 PI Korczak B, Lew A;

```

XX WPI: 2000-594746/57.
DR
XX
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
XX
PS Claim 3; Page 53; 66pp; English.
XX
XX
CC The present sequence is derived from a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
XX
SQ Sequence 406 AA:

```

```

KW Screening; treatment; prevention; cardiomyopathy; inhibitor;
KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
KW UDP-GlcNAc:Galbeta1-3GalNAc-alpha; transgenic animal; germ line;
KW beta-1,6-N-acetylglucosaminyl-1-transferase; human; core 2 GNT.
XX
XX Homo sapiens.
XX
XX CA2186987-A.
XX
XX 02-APR-1998.
XX
XX 02-OCT-1996; 96CA-2186987.
XX
XX 02-OCT-1996; 96CA-2186987.
XX
XX (MOUN ) MOUNT SINAI HOSPITAL CORP.
XX
XX Dennis JW, King GL, Koya D, Nishio Y, Warren CE.
XX
XX WPI: 1998-399608/35.
XX
XX N-PSDB; AAX24043.
XX
XX
XX This invention describes a method for screening for a substance that
XX may be used to prevent or treat cardiomyopathy associated with diabetes
XX and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
XX (UDP-GlcNAc:Galbeta1-3GalNAc-alpha-beta-1,6-N-acetylglucosaminyl-
XX transferase) with an acceptor substrate and a sugar nucleotide donor in
XX the presence of a test substance under conditions whereby the core 2
XX GlcNAc-T produces a reaction product, determining the amount of reaction
XX product, and comparing the amount of reaction product with the amount
XX obtained in the absence of the test substance, where lower amounts of
XX reaction product in the presence of the test substance indicate that the
XX substance inhibits core 2 GlcNAc-T. The invention also describes (1)
XX methods for preventing or treating cardiomyopathy associated with
XX diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
XX germ cells and somatic cells all contain a DNA construct introduced into
XX the animal or an ancestor of the animal at an embryonic stage, where
XX incorporation of the DNA construct into the germ line of the animal
XX causes the animal to develop cardiomyopathy similar to that associated
XX with diabetes mellitus and hyperglycaemia. This sequence represents the
XX human core 2 GNT protein (beta-1,6-N-acetylglucosaminyltransferase)
XX which is used to describe the method of the invention.
XX
XX
SQ Sequence 428 AA:

```


CC of activated T lymphocytes which express the C2Gnt. Transfected
CC cells were selected using MAb T305, which identifies a
CC hexasaccharide on leukostallin. Leukostallin CD43 is an acceptor
CC molecule for C2Gnt activity. A plasmid, pCDR alpha-Leu,
CC which directed expression of the T305 antigen was identified. The
CC cDNA insert was isolated and sequenced. The cDNA encoded the
CC acceptor molecule leukostallin CD43.

XX Sequence 428 AA:

Query Match 47.9%; Score 1129.5; DB 15; Length 428;
Best Local Similarity 56.9%; Pred. No. 1.1e-102;
Matches 209; Conservative 57; Mismatches 100; Indels 1; Gaps 1;

QY 68 INCSGVTGDDQEAVALINLNLEVK-KKRPFTDTHTLSLRDCEHFAERKFIOPPLSK 126
DB 57 INCTKVLGQDVNEIGKVLKLEITVCKFKRPVDDYIMTSCSSFFIKRPYIVELPSK 116
QY 127 EEVEFPFIAISWVTHEKIEENFERLLRAVAPONTICVHDEKSPETREAVKAIISCFPNV 186
DB 117 EEEEFPIAYSLVHVKIEMIDRLLSAIYIPQITFCVHVDKTSSEDSYLAAMGJASCFISNV 176
QY 187 FIASKILVRVYVYASWSRVQADLNCMEDLQSSVPKRYFLNCTGDPFIKSNAEVQALKML 246
DB 177 FVASLIESVYVASWSTVQADLNCMDLYAMSTWKYILNCGMDPIKTLNLEIVRKIL 236
QY 247 NGRNSMESEVPRKHKETRMKYHREVRDTHLTKKRRDPPRYNLTFPGNAVYASRDFV 306
DB 237 MGNMLETETMPHKEERWKYEVVNGKLTGTCTVKMLPRLTPIFSGSAYIVSREY 296
QY 307 QHVLKPKSQOQLIEWYKDTYSPDEHLMATLQARAKMPSVPHNPKYDISMTIARLVK 366
DB 297 GYVLGQEKIQKIMEWAGQDYSPEYIWTATQIPREVGPSIPASHKYDLSDMGAVARFVK 356
QY 367 OGHEGIDIGAPYAPCSGIHORAICVYAGDLNMLQNHHLANKRPKVDNALQCLE 426
DB 357 GYFEGDVSKGAPYPPCSGVHVRVSVCLFGAGDLNMLIRKHHLFANKIDVDVDFALQCLDE 416
QY 427 YLRKYAI 433
DB 417 HLTHKAL 423

RESULT 11

AAM93942
ID AAM93942 standard; Protein; 428 AA.

AC AAM93942;

XX 28-JUN-1999 (first entry)

DE Rat DH1 protein.

KW DH1: rat; screening; treatment; prevention; cardiomyopathy; inhibitor;
KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
KW UDP-GlcNAc:Galbeta1-3GalNAc-alphaHAR; transgenic animal; germ line;
KW beta-1,6-N-acetylglucosaminyl-transferase.

OS Rattus sp.

PN CA2186987-A.

XX 02-APR-1998.

PF 02-OCT-1996; 96CA-2186987.

PR 02-OCT-1996; 96CA-2186987.

XX (MOON) MOUNT SINAI HOSPITAL CORP.

PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

DR WPI: 1998-399608/35.
DR N-PSDB: AAX24042.

XX Screening for substances that prevent or treat cardiomyopathy
XX associated with diabetes and hyperglycaemia - comprises reacting
XX core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
XX in presence of test substance

XX Example 1; Fig 3A; 35pp; English.

XX This invention describes a method for screening for a substance that
XX may be used to prevent or treat cardiomyopathy associated with diabetes
XX and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
XX (UDP-GlcNAc:Galbeta1-3GalNAc-alphaHAR beta-1,6-N-acetylglucosaminyl-
XX transferase) with an acceptor substrate and a sugar nucleotide donor in
XX the presence of a test substance under conditions whereby the core 2
XX GlcNAc-T produces a reaction product, determining the amount of reaction
XX product, and comparing the amount of reaction product with the amount
XX obtained in the absence of the test substance, where lower amounts of
XX reaction product in the presence of the test substance indicate that the
XX substance inhibits core 2 GlcNAc-T. The invention also describes (1)
XX methods for preventing or treating cardiomyopathy associated with
XX diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
XX germ cells and somatic cells all contain a DNA construct introduced into
XX the animal or an ancestor of the animal at an embryonic stage, where
XX incorporation of the DNA construct into the germ line of the animal
XX causes the animal to develop cardiomyopathy similar to that associated
XX with diabetes mellitus and hyperglycaemia. This sequence represents the
XX rat DH1 protein which is used to describe the method of the invention.

XX Sequence 428 AA:

Query Match 46.9%; Score 1107.5; DB 19; Length 428;
Best Local Similarity 54.6%; Pred. No. 1.6e-100;
Matches 201; Conservative 64; Mismatches 102; Indels 1; Gaps 1;

QY 67 SINGCVTGGDDQEAVALINLNLEVKKKRPFDTHTLSLRDCEHFAERKFIOPPLSK 125
DB 56 NVNCTKVLGQDVNEIGKVLKLEITVQFKRPRTTPHNDYIMTSCSSFFIKRYIMPEIT 115
QY 126 KEEVEFPFIAISWVTHEKIEENFERLLRAVAPONTICVHDEKSPETREAVKAIISCFPN 185
DB 116 KEEVGFPIAYSLVHVKIEMIDRLLSAIYIPQITFCVHVDKTSSEDSYLAAMGJASCFIS 175
QY 126 VFIASKILVRVYVYASWSRVQADLNCMEDLQSSVPKRYFLNCTGDPFIKSNAEVQALKML 245
DB 176 VFIASGLIESVYVASWSTVQADLNCMDLYAMSTWKYILNCGMDPIKTLNLEIVRKIL 235
QY 246 NGRNSMESEVPRKHKETRMKYHREVRDTHLTKKRRDPPRYNLTFPGNAVYASRDFV 305
DB 236 FEGNSLETETMPHKEERWKYEVVNGKLTGTCTVKMLPRLTPIFSGSAYIVSREY 295
QY 306 VQHVLKPKSQOQLIEWYKDTYSPDEHLMATLQARAKMPSVPHNPKYDISMTIARLVK 365
DB 296 VGYVLGQEKIQKIMEWAGQDYSPEYIWTATQIPREVGPSIPASHKYDLSDMGAVARFVK 355
QY 366 WQHEGIDIGAPYAPCSGIHORAICVYAGDLNMLQNHHLANKRPKVDNALQCLE 425
DB 356 WGYFEGDVSKGAPYPPCSGVHVRVSVCLFGAGDLNMLIRKHHLFANKIDVDVDFALQCLDE 415
QY 426 EYLRKYAI 433
DB 416 EHLTHKAL 423

RESULT 12

AAB30297
ID AAB30297 standard; Protein; 427 AA.

AC AAB30297;

DT 12-FEB-2001 (first entry)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 08:15:46 ; Search time 19.87 Seconds

(without alignments)
538,420 Million cell updates/sec

Title: US-09-874-390-2

Perfect score: 2360
Sequence: 1 MVQMKRLCOLHYLMALGCMY.....NALQCLEXYLRKAYGTETL 438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2360	100.0	438	4	US-09-233-506-2
2	1142.5	48.4	428	1	US-07-955-041-4
3	1142.5	48.4	428	1	US-08-227-455-4
4	1142.5	48.4	428	1	US-08-472-482-4
5	1142.5	48.4	428	1	US-08-487-069-4
6	1142.5	48.4	428	4	US-09-233-506-3
7	784	33.2	400	1	US-08-118-906-14
8	784	33.2	400	1	US-08-486-196-14
9	784	33.2	400	1	US-08-488-135-14
10	784	33.2	400	2	US-08-474-065-14
11	784	33.2	400	4	US-09-233-506-4
12	447	18.9	126	1	US-08-118-906-4
13	447	18.9	126	1	US-08-486-196-4
14	447	18.9	126	1	US-08-488-135-4
15	447	18.9	126	2	US-08-474-065-4
16	426	18.1	126	1	US-08-118-906-2
17	426	18.1	126	1	US-08-486-196-2
18	426	18.1	126	1	US-08-488-135-2
19	426	18.1	126	2	US-08-474-065-2
20	278	11.8	64	4	US-09-233-506-10
21	203.5	8.6	49	4	US-09-233-506-14
22	126	5.3	33	1	US-08-118-906-8
23	126	5.3	33	1	US-08-486-196-8
24	126	5.3	33	1	US-08-488-135-8
25	126	5.3	33	2	US-08-474-065-8
26	118	5.0	33	1	US-08-118-906-6
27	118	5.0	33	1	US-08-486-196-6

28	118	5.0	33	1	US-08-488-135-6	Sequence 6, Appl1
29	118	5.0	33	2	US-08-474-065-6	Sequence 6, Appl1
30	109	4.6	22	1	US-08-118-906-12	Sequence 12, Appl1
31	109	4.6	22	1	US-08-486-196-12	Sequence 12, Appl1
32	109	4.6	22	1	US-08-488-135-12	Sequence 12, Appl1
33	109	4.6	22	2	US-08-474-065-12	Sequence 12, Appl1
34	103	4.4	316	1	US-08-597-236-12	Sequence 12, Appl1
35	103	4.4	316	1	US-08-746-682A-12	Sequence 12, Appl1
36	101	4.3	788	2	US-08-907-166-6	Sequence 36, Appl1
37	100.5	4.3	2555	3	US-09-058-489-36	Sequence 36, Appl1
38	94.5	4.0	253	3	US-08-929-329-9	Sequence 9, Appl1
39	94.5	4.0	619	4	US-09-153-310-42	Sequence 42, Appl1
40	92.5	3.9	2289	3	US-09-051-019-2	Sequence 2, Appl1
41	91	3.9	3135	1	US-08-323-170B-2	Sequence 2, Appl1
42	91	3.9	3135	4	US-08-954-441-2	Sequence 2, Appl1
43	90.5	3.8	616	4	US-08-965-902A-2	Sequence 2, Appl1
44	90	3.8	1128	1	US-08-111-939-2	Sequence 2, Appl1
45	90	3.8	1194	4	US-09-092-508-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1	US-09-233-506-2	Sequence 2, Application US/09233506
Patent No. 6136580	GENERAL INFORMATION:	
APPLICANT: Fukuda, Minoru	TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms	
TITLE OF INVENTION: Core 2, Core 4 and I Branches	FILE REFERENCE: P-1J 3415	
CURRENT APPLICATION NUMBER: US/09/233, 506	CURRENT FILING DATE: 1999-01-19	
NUMBER OF SEQ ID NOS: 14	SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO 2	LENGTH: 438	
TYPE: PRT	ORGANISM: Homo sapiens	
US-09-233-506-2	Query Match	100.0%; Score 2360; DB 4; Length 438;
	Best Local Similarity	100.0%; Pred. No. 6.9e-230; Mismatches 0; Gaps 0;
	Matches 438; Conservative 0; Indels 0;	
QY 1	MVQMKRLCOLHYLMALGCMYLMATVALKLSFRKCDSDHGLGSRSSOSQYCRNLYNPL 60	
DB 1	MVQMKRLCOLHYLMALGCMYLMATVALKLSFRKCDSDHGLGSRSSOSQYCRNLYNPL 60	
QY 61	KLPKRSINSGVTRGDOEAVLQATLNNLEVKKKREPTDTHYLSLTFDCEHFAERKFI 120	
DB 61	KLPKRSINSGVTRGDOEAVLQATLNNLEVKKKREPTDTHYLSLTFDCEHFAERKFI 120	
QY 121	QPLSKSEVERPIAVSWIHKIENFERLLRAVAPQNYICVHDEKSPETFEKAVKAI 180	
DB 121	QPLSKSEVERPIAVSWIHKIENFERLLRAVAPQNYICVHDEKSPETFEKAVKAI 180	
QY 181	SCFPNVFASLKLVRVYASWSRVQADLNCMEDLLQSSVPWKYFTLNTCGTDPPIKSNABV 240	
DB 181	SCFPNVFASLKLVRVYASWSRVQADLNCMEDLLQSSVPWKYFTLNTCGTDPPIKSNABV 240	
QY 241	QALKMLNGRNSMESVPPKHKETRWKYHFEVVRDTLHTNNKKDPPYNTLMTGNATY 300	
DB 241	QALKMLNGRNSMESVPPKHKETRWKYHFEVVRDTLHTNNKKDPPYNTLMTGNATY 300	
QY 301	ASRPFOVHVLKNPKSQOLIEVVKDTYSPDEHLMATLQARMPGSPVHPKYDISDMTSI 360	
DB 301	ASRPFOVHVLKNPKSQOLIEVVKDTYSPDEHLMATLQARMPGSPVHPKYDISDMTSI 360	
QY 361	ARLVKQGHGDDIKGAPYAPDCGIHQRAICVYGAGDLMNMLQNHHLANKFDPKVDNA 420	

```

Db      361  ARLVKQHGEGIDKCAPAPCSGHQRAICVYGAGDLNMMLQNHHLANKFDPKVDNA      420
      |||||||
Oy      421  LOCLELYLRKKAIGTEL 438
      |||||||
Db      421  LOCLELYLRKKAIGTEL 438

RESULT 2
US-07-955-041-4
; Sequence 4, Application US/07955041
; Patent No. 5360733
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUTZEN, MARTI PA
; TITLE OF INVENTION: A NOVEL BETAL-6
; TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE
; TITLE OF INVENTION: LERKOSTALIN AND A METHOD FOR CLONING PROTEINS HAVING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,041
; FILING DATE: 19921001
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-335-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-955-041-4

```

	Query Match	48.4%	Score 1142.5	DB 1	Length 426	
	Blast Local Similarity	57.5%	Pred. No. 6.3e-107			
	Matches 211	Conservative	57	Mismatches 96	Indels 1	Gaps 1
Oy	68	INCGSVTRGDDQDAVLQALAINNLEVK-KRREPETHYLSLTRCDCHFKARKRFIOPLSK	126			
		: : :		: : : :		
Dd	57	INCFVGLQDVAEIKRKKEILTVKFRKKRPMTDPDYINMSTDSSFIKRRIIVPELKL	116			
Oy	127	EERVEPIAVSMVIHHEKTEMEFERLRVVAPQNIYCYNHDEKSPFEKPKVALIISOPPV	186			
		: : : :		: : :		
Dd	117	EAAEPPIAVSIYVHHKIEMDLRLRIITYPQNFYCHVDTKSEDSLTAAVMGIASFSNV	176			
Oy	187	FIAASKLVRRVVASWSRVQADLNCMBELQSSVPWKYFLNTCTDFPIKISNAEVQALKML	246			
		: :				
Dd	177	FVASRLSESYVVASWSRVQADLNCMKQLYMASNMWKLIMLCMDPEIKTNLEIVRLKLL	236			
Oy	247	NCRNMSSESVPVRKHETIRMYAEHYVROTTLJNNKKDDPPVNLPMTEGNATVYASRDV	306			
		: : : :		: : : :		: : : :
Dd	237	MGEENULETFRMASHKEERRKKRYEVVNGULTVTGYKMLPLETLPFSSAIFVYSREY	296			
Oy	307	OHVLYKNPSQILLIEWKDYYSPDEHLMATLQRARMPGSVPHHPKYDISDMTSIARLVKM	366			

[illegible]

```

1      RESULT      3
2      US-08-227-455-4
3      : Sequence 4, Application US/08227455
4      : Patent No. 5624832
5      : GENERAL INFORMATION:
6      : APPLICANT: FUKUDA, MINORU
7      : APPLICANT: BIERHUIZEN, MARTI EA
8      : TITLE OF INVENTION: A NOVEL BETA1-6
9      : TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE
10     : TITLE OF INVENTION: LEUCOSTALIN AND A METHOD FOR CLONING PROTEINS HAVING
11     : ENZYMATIC ACTIVITY
12     : NUMBER OF SEQUENCES: 8
13     : CORRESPONDENCE ADDRESS:
14     : ADDRESSEE: CAMPBELL AND FLORES
15     : STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
16     : CITY: SAN DIEGO
17     : STATE: CALIFORNIA
18     : COUNTRY: USA
19     : ZIP: 92122
20     : COMPUTER READABLE FORM:
21     : MEDIUM TYPE: floppy disk
22     : COMPUTER: IBM PC compatible
23     : OPERATING SYSTEM: PC-DOS/MS-DOS
24     : SOFTWARE: PatentIn Release #1.0, Version #1.25
25     : CURRENT APPLICATION DATA:
26     : APPLICATION NUMBER: US/08/227,455
27     : FILING DATE: 14-APR-1994
28     : CLASSIFICATION: 435
29     : ATTORNEY/AGENT INFORMATION:
30     : NAME: CAMPBELL, CATHRYN
31     : REGISTRATION NUMBER: 31,815
32     : REFERENCE/DOCKET NUMBER: P-LJ 9957
33     : TELECOMMUNICATION INFORMATION:
34     : TELEPHONE: 619-535-9001
35     : TELEFAX: 619-535-8949
36     : INFORMATION FOR SEQ ID NO: 4:
37     : SEQUENCE CHARACTERISTICS:
38     : LENGTH: 428 amino acids
39     : TYPE: amino acid
40     : TOPOLOGY: linear
41     : MOLECULE TYPE: protein
42     : US-08-227-455-4

```

[illegible]


```

Db 227 MGENNLETERMPSHKEERKRYEYVNGKLTNTGTVMPLPLETPLFSGSAFVVSREYV 296
OY 307 QHVLNPKPSQOLIEVWKDTYSPDEHLMTLQARMMPCSPVPHPKRYDSDMTSIARLYKM 366
Db 297 GVLQNEKIQKLEMAOPTYSPDEYLMATIQRIPEVPSLPSHXYDLSDMQAVARFYKM 356
OY 367 QGHGEDIDKAPYAPCSGIIHORAICVYGAGDLNMMLOHNLHLANKFDPKVDNALQCLEE 426
Db 357 QYFEGDVSKGAPYPCDGVHVRSCIFGAGDLNMMLRKHLFANKFEDVDVDFALQCLDE 416
OY 427 YLYRYKAI 433
Db 417 HLRHAKL 423

RESULT 4
US-08-472-482-4
; Sequence 4, Application US/08472482
; Patent No. 5658778
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETAL-6
; TITLE OF INVENTION: N-ACETYLDUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMACTIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,482
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: amino acid.
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-472-482-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

OY 68 INCSEVTGDOEAVLQALINLEVA-KRREPPTDTHYSLTRDCHEFAERKFFIPLSK 126
Db 57 INCTVVLQGVNLEIQLVLEILTVFKRKRPRTPDDYINMTSDCSSFTKRREYIPEPLSK 116
OY 127 EEVEEPPIAVSWIHEKINFERLLRAVYAPONIVCVHDEKSPERFKKAVNAIISCFNV 186

```

```

Db 117 EEAEPPIAVSIYVHHKIEKMDRLRAITWPNQFVCHVDTKESDYLAAMGIAFCFSNV 176
OY 187 FIASKLVRVYVASMGRVADLNCMEDLQSSVPMKYFLNTCGTDPPIKSNAEMOVALKML 246
Db 177 FVASHLESYVYVASMGRVADLNCMDLYAMSNMXYLLNLGMDPIKTNLEIVRKILL 236
OY 247 NGRNMESEVPPKHEKTRKXHFEEVYVDTLHLTKKKODPPYNLTMFTGNAYIVASRDFV 306
Db 227 MGENNLETERMPSHKEERKRYEYVNGKLTNTGTVMPLPLETPLFSGSAFVVSREYV 296
OY 307 QHVLNPKPSQOLIEVWKDTYSPDEHLMTLQARMMPCSPVPHPKRYDSDMTSIARLYKM 366
Db 297 GVLQNEKIQKLEMAOPTYSPDEYLMATIQRIPEVPSLPSHXYDLSDMQAVARFYKM 356
OY 367 QGHGEDIDKAPYAPCSGIIHORAICVYGAGDLNMMLOHNLHLANKFDPKVDNALQCLEE 426
Db 357 QYFEGDVSKGAPYPCDGVHVRSCIFGAGDLNMMLRKHLFANKFEDVDVDFALQCLDE 416
OY 427 YLYRYKAI 433
Db 417 HLRHAKL 423

RESULT 5
US-08-487-069-4
; Sequence 4, Application US/08487069
; Patent No. 5684134
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETAL-6
; TITLE OF INVENTION: N-ACETYLDUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMACTIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,069
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-487-069-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

```

Qy	68	INCSVWRGDEAVLOALINNNLEVK-KKREPEPTDHTSLTPROENHKARKFLQFLOPSK	126
Db	57	INCKRVLQAGDVNEIQAKKLELTYKFKKRPWRPDDIYNNTSDOSSFTIKRKRTIYVEPSK	116
Qy	127	BEVEFPYATSWVIEHKETENFERLLRAYABONITCVAVDEKSPETFEAAVKAATISCPNV	186
Db	117	EEAEFPYATSIYVHNKKTEMDRLRALTYMQNFYCVAHDTKSEDSYLAAMGJASCSNV	176
Qy	187	FIASKLVRVYVYASMSRVQADLNCKMEDLLOSSVPKRYFLNTGCDPFIKSAENVOALKML	246
Db	177	FVASRLSEVYVYASMSRVQADLNCKMKDDIYAMSANKYILINCGMDFPIKTYMLEIYRKLL	236
Qy	247	NGRSMSESEVPRKKETRWYKHEFVVVBDTLNTLTKKKDPRPNLTMGTGAATYVASHDY	306
Db	237	MEBNNLETERPSPKHEERWKKRYHEVVVNGKLLNTGTVMALPPLPTPLSGSAYFVAVSEYV	296
Qy	307	QHVLTNPKSQOOLIEWVDKYTSPDEHMTATLQARFMGSPVPHNKYDVISMTSIARLVKN	366
Db	297	GTVLQNEKIQKLMEMADQYTSPEDEYMATQRIPEYVGSILPASHIKYDLSMQAVARVKN	356
Qy	367	QGHGEGIDKGPARYAPCSGIHQRAITCVYGAADLNMLLNHHLLANKFDPKYVDNALQLEF	426
Db	357	QYFEGDVSKGAPYRPPCGVHVRSYCIAGADLNMLRKHHLLFANKFPVDVDFEALQCLDE	416
Qy	427	YLRTYKAI 433	
Db	417	HLRKKAL 423	

RESULT 6
 US-09-233-506-3
 Sequence 3, Application US/09233506
 Patent No. 6136580
 GENERAL INFORMATION:
 APPLICANT: Fukuda, Minoru
 APPLICANT: Yeh, Jiunn-Chern
 TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
 TITLE OF INVENTION: Core 2, Core 4 and I Branches
 FILE REFERENCE: P-LJ 3415
 CURRENT APPLICATION NUMBER: US/09/233,506
 CURRENT FILING DATE: 1999-01-19
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 3
 LENGTH: 428
 TYPE: PRF
 ORGANISM: Homo sapiens
 US-09-233-506-3

Query Match	48.4%;	Score 1142.5;	DB 4;	Length 428;
Best Local Similarity	57.5%;	Pred. No. 6.3e-107;		
Matches 211;	Conservative 57;	Mismatches 98;	Indels 1;	Gaps 1

QY	68	INCSGVRRGDDEAVLOALINNNLEVK -KKRPEDPTDITSLTRQCEHNKAKRKRJOPLSK	126
Db	57	INCRKVLQGDVNELOQKVKLEITLVKFKKRRPRMRPDVQINNTSDOSSPKRKRKKTIVELSK	116
QY	127	EVEEPFLIATSMVTHEKTEENPERELLIRAVYAQONITCVAVDEKSEPTREKVAOKALITSCPNN	186
Db	117	EEAEPFLIATSIYVHHKLTMDRLRLATAYMQNEFCVAVHDKSDSYIAAAYGASCISNV	176
QY	187	FIASKLVRVVVYASMSRQADDLNCMEDLLOSSVPMKFLNTGCTDFPIKSAENAVOALKML	246
Db	177	FVAERLESVVYASMSRQADDLNCMKDKDYASAMNNKYLINCGMDFPIKTNLEIRKTKLL	236
QY	247	NGRSMSEVDPKPKETFRWXYHFEYVVDITLNTKKKDDPRXYMLTMTGAAYIVASRDV	306
Db	237	MGEHNMLETERPKSKERPMKKRYIVVNGKILNTGYKMLPRLPLELSSGATFYVVSREY	296
QY	307	QHVILNPKSOOLIEWVKDYSPEDEHMTLQORARMMGSGVPHNPKXDISMTSIARLVK	366

Db	2 9 7	g y v l o n e k i o k l m e a d o t s p d e y l m a t i o r i p e y g s l p a s h k y d l s d m a v a r p v k	3 5 6
Oy	3 6 7	o g h e d i d d k g a r p y c s i h o r a i c y v g a g d l m m l o n h l l k m l e r p k y d d a l o c l e e	4 2 6
Db	3 5 7	q y f e d y s k g a r p y c d s v n r s c i f g a d l m m l k h n l f a n k r e d v d y d l a i o c d e	4 1 6
Oy	4 2 7	y l r t k a i	4 3 3
Db	4 1 7	h l r h k a l	4 2 3

```

1      RESULT 7
2      US-08-118-906-14
3      ; Sequence 14, Application US/08118906
4      ; Patent No. 5484590
5      ;
6      ; GENERAL INFORMATION:
7      ; APPLICANT: Fukuda, Minoru
8      ; APPLICANT: Bietnuizen, Marti F.A.
9      ; TITLE OF INVENTION: Expression of the Developmental I
10     ; TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
11     ; TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
12     ; NUMBER OF SEQUENCES: 14
13     ; CORRESPONDENCE ADDRESS:
14     ; ADDRESSEE: Campbell and Flores
15     ; STREET: 4370 La Jolla Village Drive, Suite 700
16     ; CITY: San Diego
17     ; STATE: California
18     ; COUNTRY: USA
19     ; ZIP: 92122
20     ;
21     ; COMPUTER READABLE FORM:
22     ; MEDIUM TYPE: Floppy disk
23     ; COMPUTER: IBM PC compatible
24     ; OPERATING SYSTEM: PC-DOS/MS-DOS
25     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
26     ; CURRENT APPLICATION DATA:
27     ; APPLICATION NUMBER: US/08/118,906
28     ; FILING DATE: 09-SEP-1993
29     ; CLASSIFICATION: 536
30     ;
31     ; ATTORNEY/AGENT INFORMATION:
32     ; NAME: Campbell, Cathryn A.
33     ; REGISTRATION NUMBER: 31,815
34     ; REFERENCE/DOCKET NUMBER: P-LJ 9526
35     ; TELECOMMUNICATION INFORMATION:
36     ; TELEPHONE: (619) 535-9001
37     ; TELEFAX: (619) 535-8949
38     ; INFORMATION FOR SEQ ID NO: 14:
39     ; SEQUENCE CHARACTERISTICS:
40     ; LENGTH: 400 amino acids
41     ; TYPE: amino acid
42     ; TOPOLOGY: linear
43     ; MOLECULE TYPE: protein
44     ;
45     ; US-08-118-906-14

```

Query Match	33.2%;	Score 784;	DB 1;	Length 400;
Best Local Similarity	42.3%;	Pred. No. 9.1e-71;		
Matches 163;	Conservative 62;	Mismatches 126;	Indels 34;	Gaps 9;

Qy	56	LYNFKLPARKRINCSGVTRGDDEAVLQALINLNLEVKKKREPTDTHYLSITPDCHEFKA	115
Db	24	VFNNGGDSPFRLLISPLRLTO--VQTSIF--KTFELKNNKLMTHESKSEYELT	77
Qy	116	ERKEIOPLSKEEVEEPFIAYSMVTHEKIEENFERLRAVVPQNTICVHNDEKSPFEKKA	175
Db	78	QSHITTAPLSKEEDPEFLATIMYTHHHFDFAFLFRATIPQNTICVHNDEKATTEFKKA	137
Qy	176	VKAIISCEPNVFIASKLVRRVYVYASWSHVOADLNCMEDLLQSSVPKRYFLNTCGTDFPIKS	235
Db	138	VEQLLSCFPNAPFLASKMEPVYGGISRLQDLNLCIRDLASFVEMSKYVINTCGODEPLKT	197
Qy	236	NAEVOALKMLNGNSMESEVPRPHKTRPKRY-HFE-----VYRDILHLNKKKDDP	286
Db	198	NKEIYVQLKGEKGNITPGVLPNHALGRKRVYVHQLKELSYLR-----TALKPP	252

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 08:19:26 ; Search time 38.79 Seconds
(without alignments)
1084.999 Million cell updates/sec

Title: US-09-874-390-2

Perfect score: 2360

Sequence: 1 MYQWKRLCOLHYLMALGCMY.....NALQCLEYLRKAYGTSL 438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1142.5	48.4	428	2 A46293	beta-1,3-galactosyl
2	784	33.2	400	2 A46297	beta-1,6-N-acetylgl
3	379	16.1	472	2 T22188	hypothetical prote
4	371	15.7	445	2 T21282	hypothetical prote
5	367	15.6	401	2 T24929	hypothetical prote
6	352	14.9	425	2 T20207	hypothetical prote
7	349.5	14.8	402	2 T24930	hypothetical prote
8	344.5	14.6	454	2 T21261	hypothetical prote
9	271	11.5	470	2 T32137	hypothetical prote
10	267	11.3	467	2 T21408	hypothetical prote
11	267	11.3	483	2 T89364	hypothetical prote
12	259.5	11.0	322	2 D87603	glycosyl transfera
13	254	10.8	454	2 T25382	hypothetical prote
14	248.5	10.5	513	2 T20068	hypothetical prote
15	232	9.8	489	2 T24013	hypothetical prote
16	219	9.3	459	2 T24742	hypothetical prote
17	213.5	9.0	486	2 T21566	hypothetical prote
18	212.5	9.0	367	2 T28892	hypothetical prote
19	210.5	8.9	478	2 T27714	hypothetical prote
20	205.5	8.7	753	2 T24745	hypothetical prote
21	199	8.4	472	2 T33384	hypothetical prote
22	177	7.5	406	2 T96571	hypothetical prote
23	176	7.5	448	2 D85042	hypothetical prote
24	166	7.0	395	2 B86735	unknown protein F2
25	164.5	7.0	447	2 T00906	hypothetical prote
26	157	6.7	434	2 T51450	hypothetical prote
27	150	6.4	384	2 T08940	hypothetical prote
28	145	6.1	371	2 T21819	hypothetical prote
29	125	5.3	630	2 T02524	probable RING zinc

30	116	4.9	298	2 AD2346	hypothetical prote
31	108	4.6	1201	2 H69898	hypothetical prote
32	105	4.4	787	2 C75068	probable beta-gala
33	104.5	4.4	300	1 ND8P7	exodeoxyribonuclea
34	101.5	4.3	302	1 S07503	exodeoxyribonuclea
35	100	4.2	897	2 G89923	hypothetical prote
36	99.5	4.2	876	2 G89952	DNA polymerase I
37	99.5	4.2	946	2 T31488	hypothetical prote
38	99	4.2	690	2 C96572	protein F12M16.4
39	97	4.1	823	2 G83905	hypothetical prote
40	97	4.1	987	1 T0B530	transposase - Bacti
41	96.5	4.1	779	1 S61571	suppressor protein
42	95.5	4.0	735	2 A82817	pentaphosphate qua
43	95	4.0	932	2 T25008	hypothetical prote
44	94.5	4.0	498	2 T47919	hypothetical prote
45	94.5	4.0	671	2 S61693	probable membrane

ALIGNMENTS

RESULT 1

A46293 beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase

C:Species: Homo sapiens (man)

C:Date: 20-Oct-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000

C:Accession: A46293

R:Blernhuizen, M.F., Fukuda, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 9326-9330, 1992

A:Title: Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcN

A:Reference number: A46293; MUID:93028457

A:Accession: A46293

A:Molecule type: mRNA

A:Residues: 1-428 <BIE>

A:Cross-references: GB:M97347; NID:g183440; PIDN:AAA35919.1; PID:g183441

A:Experimental source: HL-60 cells

A:Note: sequence extracted from NCBI Backbone (NCBIW:115900, NCBIIP:115901)

C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	Best Local Similarity	Score	DB 2;	Length	428;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;					
Qy 68	INCSGVTGQDQAVLQATLNNLEVK-KKREPTDTHVSLTRDCSHFAERKFIQFPLSK 126				
Db 57	INCTVVLQDVNEIQVKLEILTVFKRPRTDPDYNMTSDCSSFKRKRYVEPLSK 116				
Qy 127	EEVEPPIAVSWVHHEIKENFERLRAVYAPQNIYCVHDEKSPETFEKAVKAIISCFPNV 186				
Db 117	EEAEPPVAVSIYVHHKIKEMDLRLAIRATMPQFVCHVDTSSEDSYLAAMVGIACFSNV 176				
Qy 187	FIASKLVRVYVYASMSRVADLNCMEDLLQSSVPWKYFPLNTGCTGTPPIKSNAMVQALMKL 246				
Db 177	FVASKLVSVYVYASMSRVADLNCMKDLYAMSNMKYLLINLCMDPPIKTINTEIVAKLL 236				
Qy 247	NGRNSMESEVPKHKETEMKRYHEVYVRDYLHNTKKKDRPPYNNLMPGNAYIVASRQV 306				
Db 237	MGNNLETERMSHKEEMKRYEYVNGKLTNTGTVKMLPLEPPLFSGSAFVVSRETV 296				
Qy 307	QHVLTNPKSOQLIEVVKDTPSPDEHLMATLQARAMPSPVNHPRKYDTSIMTSIARLWK 366				
Db 297	GYVLQNEKIQKIMENAQDTYSDEFLMTIQRIPEVPSLSASHKYDLSDMQAVARFVKW 356				
Qy 367	QGHESDIDKGAVPACSGITHORATCVGAGDLNMLQNHHLANKFDPKVDNALCLEE 426				
Db 357	QYFEEDVSGKAPYPCDGVHVRSCVIFGAGDLNMLRKHHLFANFEDVDVDFALQICDE 416				
Qy 427	YLRYKAI 433				
Db 417	HLRRKAL 423				

RESULT 3
T22188
hypothetical protein F44F4.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22188
R:Coles, L.
submitted to the EMBL Data Library, September 1994
A:Reference number: Z19528
A:Accession: T22188
A:Status: preliminary; translated from GB/EMBL/DDBT
A:Molecule type: DNA
A:Residues: 1-472 <WIL>
A:Cross-references: EMBL:Z37092; PIDN:CAA85457.1; GSPDB:GN00020; CESP:F44F4.6
A:Experimental source: clone F44F4
C:Genetics:
A:Gene: CESP:F44F4.6
A:Map position: 2
A:Introns: 57/2; 105/2; 185/3; 261/3; 319/1; 370/1; 412/3

	Query Match	15.7%	Score 371	DB 2	Length 445
	Best Local Similarity	25.5%	Pred. No. 8.5e-21		
	Matches	98	Conservative	76	Mismatches 149; Indels 62; Gaps 9
QY	68	INCSVYTGDDQEAVALQALINLEAVKKKKEPPFDTHYLSITRCCHEPFAKERTIQPPLSKE	127		
DB	62	INCLNFIENGSKNRNRDLRTVNSRSIDK-----LVYSTDRCLQLKLSIFRKNKPLASE	113		
QY	128	EEVEPIAYSMYIHEK-----IENFERLLRAVYAPONICVAVHDEKSPETFEKAVVAIIISC	182		
DB	114	EESEFLSGGLVYKELSGVTTQRTVFLMSSIIHPNCEICANGENSAPIFQULLLELSSC	173		
QY	183	FPNWFIAKLVYRAYSASRVQADLNCKEDDLQSSVPMKYFLNTCGTDEPIKSNABMQA	242		

Db 174 FSNHIF-MKRPPIIDMGSHIINSAYDCLLEFLSHLSDMRYFOYLSGVDIPLKTNLEWVOI 232
QY 243 LKMLNGSNSESEVPPKHEPRMYHEVVRDTHLTKKKDDPPYVLTJMTGNAYIVAS 302
Db 233 LKHLNGANANVEIK-PIYOYORLKGK-----NEQOSPLP-----LFKSSLSLIP 274
QY 303 RDEVOYHVLKPNKSOQLIEWKDYSPDEHLMATLORARMMGVSYPNHPKYDISDSTIAR 362
Db 275 REANHLSSSSIPQOLLEPLRNTGIDAEQFWGTL-----FGNKLFPDIPGSLN 322
QY 363 LVKMOGHEGIDKCAPAP-----CSGHORAIQYAGADLMMMLQNH 405
Db 323 EKEMISYKNVNETMLTPTDGMRYIISRDQIWSKPNCHNYMKAGSCVFGLGDVPLRLSK 382
QY 406 HLANKFDPKVDNALOC-LEEYLK 429
Db 383 ALVAHKYILKSEPPAYFCLLEHRR 407

RESULT 5

T24929

hypothetical protein T15D6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T24929

R:Dobson, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19956

A:Accession: T24929

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-401 <M12>

A:Cross-references: EMBL:283125; PIDN:CAB05620.1; GSPDB:GN00019; CESP:T15D6.2

A:Experimental source: clone T15D6

C:Genetics:

A:Gene: CESP:T15D6.2

A:Map position: 1

A:Introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3

Query Match

Best Local Similarity 15.6%; Score 367; DB 2; Length 401;

Matches 97; Conservative 74; Mismatches 143; Indels 66; Gaps 8;

QY 57 YNFKLPAKR-----SINCGVTRGDOEAVLQAILNMLEYKKRREPTDTHYLSLT 107
Db 22 YLFAKSSRRPTDLLDDLEINCTINIIQGLKNEKOLEINTKMETDKIMNSTDR----- 75
QY 108 RDCHEFAERKFIQPLSKSEVEPRPIASVNIHEKIEFERLIRAVYAPQNYCVHYDEK 167
Db 76 --CHTVASMFREFNVPLSEEARPLSLGILVYKELSGVLFMLSTIYQPOMEYCIAVGEN 133
QY 168 SPEFFKAVKAIISCFFPNVFIASKLVAVYASMSRVOADLNCMEDLLQSSVPMKYFLNTC 227
Db 134 SASFTLLLEELSDCFPN-----KRPIWGSYEIINSVYDCLFLSHLKNMKRYFOYLS 188
QY 228 GTDPIKSNMVOALKMLNGRSMSESVPRKHEKTRMYKHYFEVVRDTLHLTKKKDPP 287
Db 189 GVDIPLKTNLEWVILSLNGTANVEITVENRR-----LIGQNETESPRL 234
QY 288 YNLTMTGNAYIVASRDFVOHVLKPNKSOQLIEWKDYSPDEHLMATLQPARMM---PG 344
Db 235 ----LFKSSLSLIPKRAANYLASSSIPOQLLEFLRNTWVADEQFWGTLPGNKKGLFDVPG 290
QY 345 SVPHNPKYDISDMTISIRLVKMOGHEGIDKCAPYAPCSGTHQRAICVYAGADLMMMLQ 404
Db 291 S-----LNFEHQIMFESG-----CHNMKDGSCVFGIGDVSNLQQA 327
QY 405 HHLANKFDPKVDNALOCL 424
Db 328 KALVAHKYILKSEPPAYFCL 347

RESULT 6

T20207

hypothetical protein C54C8.11 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C/Accession: T20207; T27478

R:Dobson, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19237

A:Accession: T20207

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-425 <M12>

A:Cross-references: EMBL:283102; PIDN:CAB05469.1; GSPDB:GN00019; CESP:C54C8.11

A:Experimental source: clone C54C8

R:McLay, K.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20372

A:Accession: T27478

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-425 <M12>

A:Cross-references: EMBL:AL032664; PIDN:CAA21769.1; GSPDB:GN00019; CESP:C54C8.11

A:Experimental source: clone Y91F4A

C:Genetics:

A:Gene: CESP:C54C8.11

A:Map position: 1

A:Introns: 29/2; 117/3; 161/3; 195/3; 237/3; 266/3; 295/1; 313/3; 355/1; 374/2

Query Match

Best Local Similarity 14.9%; Score 352; DB 2; Length 425;

Matches 111; Conservative 58; Mismatches 167; Indels 58; Gaps 14;

QY 68 INCSGVTRGDOEAVLQAILNMLEYKKRREPTDTHYLS--LTRCEHFAERKFIQPLPS 125
Db 41 LECOKMFDGPPVSLARKALFKFD--DREIIAQIILKSGOENAOCAEFQKIFGFOEPTS 97
QY 126 KEEVEPRPIASVNIHEKIEFERLIRAVYAPQNYCVHYDEKSEPFKEAVKAIISCSPN 185
Db 98 QEELEFPLAYMVLHGDFVQLSLISAIYQPONFCLAVDGNSSVEFLGLVRLMSRYGN 157
QY 186 V--FIASKLVAVYASMSRVOADLNCMEDLLQSSVPMKYFLNTCGTDFPIKSNMVOAL 243
Db 158 IQYFYTDE--IRMGVEIILTSVFCVDYLAKLPDSMKYFYGLSGVDAPLKSINLEMRIL 214
QY 244 KMLNGRSMSESVPRKHEKTRMYKHYFEVVRDTLHLTKKKDPPYVLTMTGNAYIVASR 303
Db 215 KALNG--SFNAELIP-----FEFYR-----LNKR--PMSSPLPYXTSLSATFSR 256
QY 304 DFVOHVLKPNKSOQLIEWKDYSPDEHLMATLO-----RARMPGSVPN---HPK 351
Db 257 KSNAPVAVSEKVLQIDPLRGTTCADSLMATIAGNPKVYKFSELPMHGGFDKAMVHKH 316
QY 352 YDISDMTISIRLVKMOGHEGIDK-----APYAPCSGTHQRAICVYAGADLNLN 399
Db 317 Y-----RRTRGKLGKY--GENCKIDNGYVSRQYVNNAP--VKCGYYRRLSCVGVVDLP 370
QY 400 WMLONHHLANKFDPKVDNALOCLEEYLKAI 433
Db 371 NLINRHELVAHKLYFSYQPAFMCLVENSROKSM 404

RESULT 7
T24930
hypothetical protein T15D6.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T24930
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19956
A:Accession: T24930

Db 271 KLFNDKSNKYDELEISGVSVHSSLSRAAVKWLVEVLSIFIDOMNODYCADQOIFIST 330
Oy 336 LORA--RMGPSVPHRPYDISDMTSIARLVKMOGHEGIDKGAAPACSGIHORAIYV 393
Db 331 FOMSDLMKPMGHFTNECHNDTAITISRLARWPIY--LDAHAHCA--TGIVRHDICTIF 386
Oy 394 GAGDLNMWLNHHLANKFD-----PKYDNDALOCLEEYLRYKAIY 435
Db 387 GIEDER-----AISKEPILNMLPARDYSIETCTAELLHNRTYLG 425

RESULT 10
T21408
hypothetical protein F26D2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T21408
R:McMurray, A.
submitted to the EMBL Data Library, November 1996
A:Accession number: Z19418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-467 <WIL>
A:Cross-references: EMBL:Z81513; PIDN:CA04185.2; GSPDB:GN00023; CESP:F26D2.3
A:Experimental source: clone F26D2
C:Genetics:
A:Gene: CESP:F26D2.3
A:Map position: 5
A:Introns: 39/1; 90/3; 194/1; 281/1; 320/2; 356/3; 402/1; 438/3
C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 11.3%; Score 267; DB 2; Length 467;
Best Local Similarity 24.8%; Pred. No. 9.1e-13;
Matches 102; Conservative 72; Mismatches 176; Indels 62; Gaps 17;

Oy 42 LESRSQSOYCRNIIYNFLKLPARKSINGVTRGDOEAVLOAILNNEVKKKREPFTDT 101
Db 55 IEALRMNDKYCKDRRF-YCKRPETHVDCRRVLRGD-KAYLOSITG-----INRIPLIEN 107
Oy 102 HYLSTRCCEHFKARKEFTQPLSKSEVEEFPYASVMVIEKLENFERLLRAYAPQNTYC 161
Db 108 PFLNLTCSAIKRRITIPKSSQFKLMLN---GTAFAFIYADVEFEKQOASVHPQNFEC 164
Oy 162 VAVDEKSPTEKEAVKAIISCPPNVFIASKLVRYVYASRSVQADLNCMEDILQSSVPEK 221
Db 165 FAVDANSSAEFQKRMKALERCLPNVFVLPYESYDSKGNINIMAHYNCMK-RLEASRGWG 223
Oy 222 YFLNTCGTDFPIKSNAEVQALKMLNGRNS--MESEVP-PKHKEPRMRYHFEVADTLHL 278
Db 224 YIMLQNHVDVITKSYELDRIFDLAGANDVPMKSEIPERRKKHKSMDL-----KSLRL 277
Oy 279 TNKKKDPPEYNTLMTFTGNAYIVASRDFVOHVLKNPKSOQLIEWVD-----TY 326
Db 278 FPNESRNSDYELEISG-----FVQASL-----SRAYVEMLVNEDVSTILQOPEY 324
Oy 327 SPDEHLMATLQAR--WMPGSVPHRPKYDISMTSIARLVKMOGHEGIDKGAAPACSG 384
Db 335 GADQFISTFQSNQDFKPMGHFTNECFH--NDTSKIFRESRWV--PINLERSNCA--SG 378
Oy 385 IHORAIQYVAGADLNMWLNHHLANKFDP-KVDNDALOCLEEYLRYKAIY 435
Db 379 TYRHDICTIFGIDF-----LAVSKYPTLTDFDSIETCTAELLHNRTYLG 422

RESULT 11
F89364
protein F26D2.3 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: F89364

Ranonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating b1o
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: F89364
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:chr_V; PIDN:CA04185.1; PID:93924744; GSPDB:GN00023; CESP:F26D2.3
C:Genetics:
A:Gene: F26D2.3
A:Map position: 5
C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 11.3%; Score 267; DB 2; Length 483;
Best Local Similarity 24.8%; Pred. No. 9.6e-13;
Matches 102; Conservative 72; Mismatches 176; Indels 62; Gaps 17;

Oy 42 LESRSQSOYCRNIIYNFLKLPARKSINGVTRGDOEAVLOAILNNEVKKKREPFTDT 101
Db 71 IEALRMNDKYCKDRRF-YCKRPETHVDCRRVLRGD-KAYLOSITG-----INRIPLIEN 123
Oy 102 HYLSTRCCEHFKARKEFTQPLSKSEVEEFPYASVMVIEKLENFERLLRAYAPQNTYC 161
Db 124 PFLNLTCSAIKRRITIPKSSQFKLMLN---GTAFAFIYADVEFEKQOASVHPQNFEC 180
Oy 162 VAVDEKSPTEKEAVKAIISCPPNVFIASKLVRYVYASRSVQADLNCMEDILQSSVPEK 221
Db 181 FAVDANSSAEFQKRMKALERCLPNVFVLPYESYDSKGNINIMAHYNCMK-RLEASRGWG 239
Oy 222 YFLNTCGTDFPIKSNAEVQALKMLNGRNS--MESEVP-PKHKEPRMRYHFEVADTLHL 278
Db 240 YIMLQNHVDVITKSYELDRIFDLAGANDVPMKSEIPERRKKHKSMDL-----KSLRL 293
Oy 279 TNKKKDPPEYNTLMTFTGNAYIVASRDFVOHVLKNPKSOQLIEWVD-----TY 326
Db 294 FPNESRNSDYELEISG-----FVQASL-----SRAYVEMLVNEDVSTILQOPEY 340
Oy 327 SPDEHLMATLQAR--WMPGSVPHRPKYDISMTSIARLVKMOGHEGIDKGAAPACSG 384
Db 341 GADQFISTFQSNQDFKPMGHFTNECFH--NDTSKIFRESRWV--PINLERSNCA--SG 394
Oy 385 IHORAIQYVAGADLNMWLNHHLANKFDP-KVDNDALOCLEEYLRYKAIY 435
Db 395 TYRHDICTIFGIDF-----LAVSKYPTLTDFDSIETCTAELLHNRTYLG 438

RESULT 12
D87603
glycosyl transferase, probable CC2860 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: D87603
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Neilson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Ko n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87603
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <STO>
A:Cross-references: GB:AE005673; NID:913424474; PIDN:AAK24824.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2860

Query Match 11.0%; Score 259.5; DB 2; Length 322;
Best Local Similarity 26.1%; Pred. No. 2.1e-12;

	Matches	80;	Conservative	53;	Mismatches	131;	Indels	43;	Gaps	10;	
Qy	133	IAYSNIHEKIE	NERELLRAVYA	PONICV	YHDEKSP	EFKEA	YAKAIISCF	PWFIAS	LU	192	
		:: :	:: :	:: :	:: :	:: :	:: :	:: :			
Db	27	IAYLVHRYPAQ	FRLEFRALHD	PNNYLV	YHDKNS	GPALQAEIR	DFLEFLA	YNAVL	ES	85	
Qy	193	VRVVYASMSR	VAQDNCMED	LLOSSV	PKYELF	INTCGT	DEPRIS	MAEMOAL	KMLN	252	
		::: :	::: :	:: :	:: :	:: :	:: :	:: :			
Db	86	KKALMGGS	LVDAELRG	MTLEMG	RDMEF	INLSGQD	EPFLYQ	KIRAF	LQNR	145	
Qy	253	-----E	SEPPKHE	KETRM	KHYEV	VR-----	DTLH	TNKK	KDP	302	
		:: :	:: :	:: :	:: :	:: :	:: :	:: :			
Db	146	RVLDDARR	RRP-----	DTMG	RVLQ	HVELK	GRIV	DTL--	VT	197	
Qy	303	RDFVQVH	LKNPKSQ	OLIEW	YKDY	SPDEH	LMATL	ORARR	MPGS	362	
		:: :	:: :	:: :	:: :	:: :	:: :	:: :			
Db	198	RAFCD	FVCH	HDPS	VDRY	KAFYR	NTFL	ADEG	FQ	247	
Qy	363	LYKMG	CHED	DDK	GAP	YAP	CSG	IHQ	ALCV	422	
		::: :	::: :	::: :	::: :	::: :	::: :	::: :	::: :		
Db	248	LIDMT	-	PGDD	K-----	LRPR--	-	TV	AA	293	
Qy	423	CLEEY	LR	429							
		:: :	:: :	:: :	:: :	:: :	:: :	:: :			
Db	294	LLEA	HLR	300							

RESULT 13
T25382
hypothetical protein T27F6.1 - *Caenorhabditis elegans*
C|Species: *Caenorhabditis elegans*
C|Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 20-Jun-2000
C|Accession: T25382
R|Dodson, R.
submitted to the EMBL Data Library, November 1996
A|Reference number: Z20026
A|Accession: T25382
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-454 <TILL>
A|Cross-references: EMBL:Z82060; PIDD:CA804881.1; GSPDB:GN00019; CESP:T27F6.1
A|Experimental source: clone T27F6
C|Genetics:
A|Gene: CESP:T27F6.1
A|Map position: 1
A|Intons: 80/3; 191/1; 278/1; 325/3; 365/2; 407/3
C|Superfamily: *Caenorhabditis elegans* hypothetical protein H4IC03.3

Query Match	10.8%	Score 254	DB 2	Length 454
Best Local Similarity	21.8%	Pred. No. 8.8e-12		
Matches 103; Conservative	80;	Mismatches 196;	Indels 94;	Gaps. 19

[illegible]

```
QY 283 KDPPEYUJEMTGNAYIASDEFOJHNLKPKSOQLLEWY-----KDTYSP 328
Db 281 IDPILINTLT-----KFAKGAVOSSLSAAADMMFTFVLDLTFTFDIMNHGTYGV 329
QY 329 DEHLMATLORARMMGVSYPNH-----PKYDISMTSIARLVKMGQHEGIDKGAYPAPC 382
Db 330 DEOFTQAFOISDFLT--GPHGHFTDTCKIKKGLTYT--GTPRPAOMT--HGDOOSKCA----- 378
QY 383 SGIHORACVYGAGDILNMMLONHHLNANKPEPKVDNALOCLLELYRYKAIYG 435
Db 379 SKSRHGICICMGIEHLSMAKSEHLMKNKULPFDYIIIECTAEILFNRRFLG 431
```

RESULT 14
 T20068
 hypothetical protein T28F3.9 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #extl-change 20-Jun-2000
 C:Accession: T20068; T25426
 R:Mortimore, B.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: 219218
 A:Accession: T20068
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-513 <W12>
 A:Cross-references: EMBL:Z82264; PIDN:CAB05168.1; GSPDB:GN00022; CESP:T28F3.9
 A:Experimental source: clone C49C3
 R:Mortimore, B.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: 220032
 A:Accession: T25426
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-513 <W12>
 A:Cross-references: EMBL:Z82285; PIDN:CAB05303.1; GSPDB:GN00022; CESP:T28F3.9
 C:Genetics:
 A:Gene: CESP:T28F3.9
 A:Map position: 4
 A:Introns: 39/3; 76/1; 124/3; 228/1; 315/1; 362/3; 402/2; 444/3; 484/3
 ;Superfamily: *Caenorhabditis elegans* hypothetical protein H41C03.3

Query Match	10.5%	Score 248.5;	DB 2;	Length 513;
Best Local Similarity	22.9%;	Pred. No. 2.8e-11;		
Matches 91;	Conservative 75;	Mismatches 164;	Indels 67;	Gaps 18

QY	59	FLIKPARKSINC5VTDGDAVLOALINNEVKKKREPPDTHYSLEDCHEPFAERK	118
Db	105	YIKREPTOHVDCGRILNMGDK-YLOTVSN-----NRIPLIKSYLWMS-CSS-----	150
QY	119	FIOEPLSKEEVEF-----PLAYSMVTHEKIEPNERILLRAVYAPONTUYCVHDEKSPET	172
Db	151	-LQSIQIDQEGNFEETPKSGSAVAFARIYEDYELEKQVQWMSHDVNFCEVIDTKSKEF	209
QY	173	KEAVKAIISCPEPNVFIASKLVRVYVYASMSRYOADLNCMEDLOSSVPWKFLWTCGTDP	222
Db	210	KNNKKLDDCEFTIYVPAEHSFSSGHNONIAGHTECQOSILOFE-NWTYLLILLONHDVI	268
QY	223	IKSNAEMQALMKLNGRSMF--SEVPRKH-KETR-M-KYFEVYRDYLHLTNKKKDPRPY	288
Db	269	SKSYVELARIEYLIGANDVNIAGIEIDORRYPRGLKWPQAKMLKLFERNESGIDDELKRP-	326
QY	289	NLUMFTGAAUYVASRDPYQHNLKPKSQOLIEWVKDT-----YSPDEHMA	334
Db	327	---WKIASGVQAS-----LSRAAEWMLINIVDLTLFLNDFRKTYGGDEOLIS	372
QY	335	TLQ--RARW-MPGSVPNH-PKVIDSDMTSIALVYMOQHEDIDKGAARYAPCSITHORACI	391
Db	373	SFOVNSQPEFMGNHTDECSKRYGH-QOISINQWMSG--GDVYRCA-----SKTIRGIC	424
QY	392	VYAGADLNMWLLONHHLLANKFPRKVDNALQCLSEYVL	428

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 09:45:11 : Search time 19.6 Seconds

(without alignments)
865.263 Million cell updates/sec

Title: US-09-874-390-2

Perfect score: 2360

Sequence: 1 MYQWKRLLCQLHYLWALGICYM.....NALQCLEYLRYKAYGTGL 438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1142.5	48.4	428 1	GGNT_HUMAN
2	1111.5	47.1	428 1	GGNT_MOUSE
3	1106	46.9	427 1	GGNT_BOVIN
4	801.5	34.0	400 1	BGIB_MOUSE
5	784	33.2	400 1	BGIB_HUMAN
6	104.5	4.4	300 1	EXRN_BPT7
7	104.5	4.4	560 1	INRL_SHEEP
8	101.5	4.3	302 1	EXRN_BPT3
9	100.5	4.3	2555 1	FAFY_HUMAN
10	100	4.2	832 1	TRNL_CANAL
11	97	4.1	987 1	TRPA_BACTU
12	96.5	4.0	779 1	MSS4_YEAST
13	94.5	4.0	731 1	UR87_HYSA
14	94	4.0	411 1	THBG_BOVIN
15	93.5	4.0	879 1	RA50_PYRO
16	92.5	3.9	619 1	LEU1_YEAST
17	92	3.9	992 1	PHS2_DICDI
18	91.5	3.9	738 1	CHSY_USTMA
19	91.5	3.9	960 1	KIT_CHICK
20	91	3.9	2663 1	CENE_HUMAN
21	91	3.9	3135 1	S230_PLAFO
22	90.5	3.8	876 1	RPOB_NPVAC
23	90.5	3.8	1013 1	CHSA_EMENT
24	90.5	3.8	1018 1	YKC2_CAEEL
25	90.5	3.8	1230 1	UBPF_YEAST
26	90.5	3.8	2559 1	FAFX_MOUSE
27	90	3.8	669 1	MYBE_AVILE
28	90	3.8	1514 1	YD34_YEAST
29	89.5	3.8	354 1	YSCU_YERPE
30	89.5	3.8	1597 1	GMFL_STROD
31	89.5	3.8	4969 1	RRR2_RABIT
32	89	3.8	1717 1	RRP1_HUMAN
33	89	3.8	3744 1	YHP9_YEAST

34	89	3.8	6669 1	NEBU_HUMAN	P20929 homo sapien
35	88.5	3.8	515 1	CP51_PENT	Q12664 penicillium
36	88.5	3.8	3068 1	POIG_PEMVC	001500 p genome po
37	88.5	3.8	5430 1	ACF7_HUMAN	09upn3 homo sapien
38	88	3.7	578 1	YWRJ_CAEEL	Q20471 caenorhabdi
39	88	3.7	601 1	YBM4_SCHPO	Q10338 schizosacch
40	88	3.7	996 1	PIP2_YEAST	P52960 saccharomyc
41	87.5	3.7	522 1	N162_HUMAN	P37198 homo sapien
42	87.5	3.7	540 1	AXR1_ARATH	P42744 arabidopsis
43	87	3.7	1011 1	VPS3_YEAST	P23643 saccharomyc
44	86.5	3.7	359 1	DNL1_BPT7	P00969 bacterioph
45	86.5	3.7	824 1	DPOL_METVO	P52025 methanococc

ALIGNMENTS

```

RESULT 1
ID      G6NT_HUMAN      STANDARD:      PRT:      428 AA.
AC      002742:
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-
DE      acetylglucosaminyltransferase (EC 2.4.1.102) (Core 2 branching enzyme)
DE      (Core2-GlcNAc-transferase) (C2GNT) (Core 2 GNT).
CN      GCM1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93028457; PubMed=1329093;
RA      Bierhuizen M.F.A., Fukuda M.;
RT      "Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta
RT      1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene
RT      transfer into CHO cells expressing polyoma large tumor antigen.";
RL      Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96078409; PubMed=759796;
RA      Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT      "Genomic organization of core 2 and I branching beta-1,6-N-
RT      acetylglucosaminyltransferases. Implication for evolution of the
RT      beta-1,6-N-acetylglucosaminyltransferase gene family.";
RL      Glycobiology 5:417-425(1995).
CC      -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
CC      -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-
CC      galactosyl-1,3-(N-acetyl-beta-D-glucosaminyl-1,6)-N-acetyl-D-
CC      galactosyl-1,3-(N-acetyl-beta-D-glucosaminyl-1,6)-N-acetyl-D-
CC      galactosaminyl-R.
CC      -1- PATHWAY: GLYCOSYLATION.
CC      -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC      -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES
CC      AND WEIOLD CELLS.
CC      -1- SIMILARITY: TO I-BRANCHING ENZYME (GNT).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M97347; AAA5919.1; -.
CC      EMBL: L41415; AAA96661.1; -.
CC      PIR: A46293; A46293.
CC      MIM: 600391; -.
CC      InterPro: IPR003406; Branch.

```


Db 356 WOYFEGDVSNGAPRPPCSGVHRSVCVFGAGDLSWMLRQHHLFANKFDMDDVPFAIQCD 415
 QY 426 EYLRYKATGTE 437
 Db 416 EHLRRKALENLE 427

RESULT 3

GENT_BOVIN

ID G6NT_BOVIN STANDARD: PRT: 427 AA.

AC 092180.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.102) (core 2 branching enzyme) (CORE2-GLCNAC-transferase) (C2GNT) (core 2 GNT).
 GN GCNT1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
 NC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Li C.M., Cheng P.W.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.

CC -I- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
 CC -I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-galactosyl-1,3-N-acetyl-D-galactosaminyl-R = UDP + beta-D-galactosyl-1,3-(N-acetyl-beta-D-glucosaminyl-1,6)-N-acetyl-D-galactosaminyl-R.
 CC -I- PATHWAY: GLYCOSYLATION.
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
 CC -I- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: U41320: AAA83244.1: -
 DR InterPro: IPR003406: Branch.
 DR Pfam: PF02485: Branch: 1.
 KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor; Golgi stack; Glycoprotein.
 FT DOMAIN 1 9
 FT TRANSMEM 10 32
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 33 427
 FT CARBOHD 52 52
 FT CAROHD 58 58
 FT CAROHD 95 95
 FT SEQUENCE 427 AA: 49733 MW: 4189CBFD948D4196 CRC64:

Query Match Best Local Similarity 55.4%; Score 1106; DB 1; Length 427; Matches 204; Conservative 64; Mismatches 96; Indels 2; Gaps 2;

QY 67 SINGSGVTRGDQEAVALQALINMLEVK-KRREPDTHTLSLRDCEHFAKRRKFTQPLS 125
 Db 56 NINCKRIILQGDVDEIQKVLKLESLTVKFKRARTNVDYINMGDCASFIKKRYITEPLS 115
 QY 126 KEVEFEPLTASVNIHEKTEIFRLRAYAAPONICVHVDSESPTEKFAVAIIISCPN 185
 Db 116 KEAGEPLTASVIVHKKIEMDLRLRLRAYIPQNFYCIHVDASSEKSFLLAAVGIASCFSN 175

QY 186 VEIASKLVRYVYASNSRVOADLNCMEDLLQSSVPMKRYPLNTGDTFPIKSNAMVQALKM 245
 Db 176 VFAASQLESVYVYASNSRVOADLNCMODLYQNMAGMKYLLNLGCMDFPITNLEIYRKLL 235
 QY 246 LNGRNSMESEVPPKHKETRWKTHFEVYRDTLLTLNKKDPPRYNLTFPGNAYIASRDF 305
 Db 236 LMGENNLETERKMPSHKKEKRWKHYEVVNGKLTNMGTDKIHPLERPLFGSAHFVVSREY 295
 QY 306 VOHVILKNPKSQQLIEVKTYSPEDEHMTATLQARHMPGVSVEHNHPPYDIDSMTIARLYK 365
 Db 296 VEIVLQNONIQFMEAKDTYSPDELTMTATQRIEVEPSLSLSTKYDTSMDQALARYK 355
 QY 366 WQGHREDIDKGPAPYPCSGIHQRAICVYGAGDLSWMLRQHHLFANKFDPKVDNALQCLCE 425
 Db 356 WOYFEGDVSNGAPRPPCS-VHRSVCVFGAGDLSWMLRQHHLFANKFDPDIDLPFAIQCD 414
 QY 426 EYLRYKAT 433
 Db 415 EHLRRKAL 422

RESULT 4

BETB_MOUSE

ID BGIB_MOUSE STANDARD: PRT: 400 AA.

AC P97402;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE N-acetylglucosaminide beta-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.150) (N-acetylglucosaminyltransferase) (I-branching enzyme) (IGNT) (large I antigen-forming beta-1,6-N-acetylglucosaminyltransferase).
 GN GCNT2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97280061: PubMed=9134435;
 RA Magnet A.D., Fukuda M.;
 RL "Expression of the large I antigen forming beta-1,6-N-acetylglucosaminyltransferase in various tissues of adult mice."; Glycobiology 7:285-295(1997).
 CC -I- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR BRANCHED POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
 CC -I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-galactosyl-1,4-N-acetyl-D-glucosaminyl-R = UDP + N-acetyl-D-glucosaminyl-1,6-beta-D-galactosyl-1,4-N-acetyl-D-glucosaminyl-R.
 CC -I- PATHWAY: GLYCOSYLATION.
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
 CC -I- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: U68182: AAB39621.1: -
 DR MGD: MGI:1100870: Gcnt2.
 DR InterPro: IPR003406: Branch.
 DR Pfam: PF02485: Branch: 1.
 KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor; Golgi stack; Glycoprotein.
 FT DOMAIN 1 6
 FT TRANSMEM 7 25
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 26 400
 FT LUMENAL, CATALYTIC (POTENTIAL).

Query Match 34.0%; Score 801.5; DB 1; Length 400;
 Best Local Similarity 43.2%; Pred. No. 6e-54;
 Matches 162; Conservative 59; Mismatches 129; Indels 25; Gaps 5;

FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 400 AA; 45507 MW; 193FF3882E3B9691 CRC64;

Query Match 34.0%; Score 801.5; DB 1; Length 400;
 Best Local Similarity 43.2%; Pred. No. 6e-54;
 Matches 162; Conservative 59; Mismatches 129; Indels 25; Gaps 5;

QY 76 GDOEAVLQALNNLEVKKKREPPTDHYLSLTRD-----CEHKAERKFIQFPLSK 126
 DB 29 GDSYQKLNISDSVRLSOVCSFFIDGKSRLFLWRKMLIHEKPSCTEVYQSHYITAPLSQ 88
 QY 127 EEEVFPYASVNIHEKIEENRELLRAYAPONTICVHDEKSPETFEAVAKISCPNV 186
 DB 89 EEGFPLAYVAVIHNFDTARLFRAIFPMQNTICVHDEKATREFGAVQOLVSCFPNV 148
 QY 187 FIASKLVVVYASWSRVADLNCMEDLQSSVPKRYFLNTGTDPIKSNAEVQALKML 246
 DB 149 FIASKMPEVYVGGISRLQADLNCIKIDLTSEVPKRYAINTCGQDFPIKTKKEIYQVLKGL 208
 QY 247 NGRSMSEVPPKKETRWKY-HFEVVRDYLHL---TNKKKDPPIVULMTGNAIYAS 302
 DB 209 KKKLPPGCVLPRAHAICRTYRHEHLSKELSYIRFTALKPPPHMLTTFGSAVYALS 268
 QY 303 RDEVOHVLKPKSQQLIEVWKDYSPDEHMLATLQARAWMGSPVNPVKYDISMTSIAR 362
 DB 269 REFANFPLRPRAVDLHWSKDFSPDEHFWVLTNRIPGVGSPN-----ASWTGNLR 322
 QY 363 LVKQGHGEOIDGAPAPCSGTHQRAICVYAGADLWMLQNHHLANKFDPKVDNALQ 422
 DB 323 AVKK-----MDMKAKHGCGQHVHIGICIGMDLQMLINSOLFANKFELNTYPLTVE 376
 QY 423 CLEERYLKRYAIGTE 437
 DB 377 CLEELRLERLINOSE 391

RESULT 5
 BGIB_HUMAN
 ID BGIB_HUMAN STANDARD; PRT; 400 AA.
 AC 006430;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE N-acetylglucosaminide beta-1,6-N-acetylglucosaminyltransferase
 DE (EC 2.4.1.150) (N-acetylglucosaminyltransferase) (I-branching enzyme)
 DE (IGNT).
 GN GCNT2 OR IT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93194065; PubMed=8449405;
 RA Bierhuizen M.F.A., Mattei M.-G., Fukuda M.;
 RT "Expression of the developmental I antigen by a cloned human cDNA
 RT encoding a member of a beta-1,6-N-acetylglucosaminyltransferase gene
 RT family.";
 RL Genes Dev. 7:468-478(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta; PubMed=7579796;
 RX MEDLINE=96078409; PubMed=7579796;
 RA Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
 RT "Genomic organization of core 2 and I branching beta-1,6-N-
 RT acetylglucosaminyltransferases. Implication for evolution of the
 RT beta-1,6-N-acetylglucosaminyltransferase gene family.";
 RL Glycobiology 5:417-425(1995).
 RN [3]

RP TISSUE SPECIFICITY.
 RX MEDLINE=98070745; PubMed=9405606;
 RA Sasaki K., Kureta-Miura K., Ujita M., Angata K., Nakagawa S.,
 RA Sekine S., Nishit T., Fukuda M.;
 RT "Expression cloning of cDNA encoding a human beta-1,3-N-
 RT acetylglucosaminyltransferase that is essential for poly-N-
 RT acetylglucosamine synthesis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14294-14299(1997).
 CC -I- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR BRANCHED
 CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
 CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED
 CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
 CC -I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-
 CC galactosyl-1,4-N-acetyl-D-glucosaminyl-R = UDP + N-acetyl-beta-D-
 CC glucosaminyl-1,6-beta-D-galactosyl-1,4-N-acetyl-D-glucosaminyl-R.
 CC -I- PATHWAY: GLYCOSYLATION.
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
 CC -I- TISSUE SPECIFICITY: IN THE ADULT, HIGHLY EXPRESSED IN PROSTATE AND
 CC TO A LESSER EXTENT IN SMALL INTESTINE AND COLON. BARELY DETECTED
 CC IN HEART, BRAIN, KIDNEY AND PANCREAS. NO EXPRESSION IN PLACENTA,
 CC LUNG, LIVER, SKELETAL MUSCLE, SPLEEN, THYMUS, TESTIS, OVARY AND
 CC PERIPHERAL BLOOD LEUKOCYTES. IN FETUS, HIGHLY EXPRESSED IN BRAIN
 CC AND TO A LESSER EXTENT IN LUNG AND KIDNEY. BARELY DETECTED IN
 CC LIVER.
 CC -I- DEVELOPMENTAL STAGE: ITS EXPRESSION INCREASES DRAMATICALLY DURING
 CC DEVELOPMENT AND ONCOGENESIS.
 CC -I- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: I19659; AAA81777.1; -
 DR EMBL: J19550; CAAT9610.1; -
 DR EMBL: I41607; AAA99832.1; -
 DR EMBL: I41605; AAA99832.1; JOINED.
 DR EMBL: I41606; AAA99832.1; JOINED.
 DR PIR: A46297; A46297.
 DR MIM: 600429; -
 DR MIM: 110800; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KW Golgi stack; Glycoprotein.
 FT DOMAIN 1 6
 FT TRANSMEM 7 25
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 26 400
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 400 AA; 45854 MW; 1469628690A1F43B CRC64;

Query Match 33.2%; Score 784; DB 1; Length 400;
 Best Local Similarity 42.3%; Pred. No. 1.3e-52;
 Matches 163; Conservative 62; Mismatches 126; Indels 34; Gaps 9;

QY 56 LYNFLKLPARKSINGSGVTRGDOEAVLQALNNLEVKKKREPPTDHYLSLRDCEHFA 115
 DB 24 VFNNGDPSFORLINSPLRLTQ--VCTSPING-----KTRFLKMKMLIHEKSSKEYLT 77
 QY 116 ERKFIQPLSKKEVEPPYASVNIHEKIEENRELLRAYAPONTICVHDEKSPETFEKA 175
 DB 78 QSHYITAPLSKEADPLAYVAVIHNFDTARLFRAIFPMQNTICVHDEKATTEFKDA 137
 QY 176 VKAIIISCFPNVFIASKVRVVYASWSRVADLNCMEDLQSSVPKRYFLNTGTDPIKRS 235
 I: ::::: 1:1111: 111 11:111111: 11 111:1111 11111:

```

Db 138 VEOILSCFPNNAFLASKMPVYVGGISRLQADINCRDLSAFREVSXKYVINTCGDPPLEKT 197
QY 236 NAEWYQALKMLNGRNSMSEVPPKHETRWXY-HFE-----VVRDTLHLTNKKDDP 286
Db 198 NKEIYQYLKGRKKKITGVLPRAHAIGRTKYVHOEHLKELSYIR-----TLAKPRP 252
QY 287 PYNLTMEFGNAVYASRDPVOHVLPKPSQOLIEVVKDTYSPDEHLMATLQARMMPGSV 346
Db 253 PHNLTIFYGSAVVALSRFANFVLHDPRAVDLQWMSKDTFSPDEHFVWTLNRPVPGSM 312
QY 347 PNHPRYDLSDMTSTARLYKMGQHEGIDKGAAPYACSGIHOATLCVYAGDLMMLQNH 406
Db 313 PN-----ASWTGNLRATIKWSDME--DR---HGCGHGHVYHGTICYNGDLKMLVNSPS 360
QY 407 LLANKDEPKVDNALQCLEEYLRYK 431
Db 361 LEANKEELNTPYLTVECLE--LRHR 383

RESULT 6
EXRN_BPT7 STANDARD: PRT: 300 AA.
AC P00638:
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Exodeoxyribonuclease (EC 3.1.11.3) (Exonuclease).
GN 6.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like phages.
OX NCBI_TaxID=10760;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RA Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RL locations of T7 genetic elements.";
RL J. Mol. Biol. 166:477-535(1983).
RN [2]
RP REVISION.
RX MEDLINE=90133923; PubMed=2614843;
RA Dunn J.J.;
RL Unpublished results, cited by:
RL Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RL J. Mol. Biol. 210:687-701(1989).
CC -1- FUNCTION: This enzyme is essential for phage DNA replication; it
CC is believed to function in the removal of DNA-linked RNA primers.
CC It is also necessary for host DNA degradation and phage genetic
CC recombination.
CC -1- CATALYTIC ACTIVITY: Degradation of double-stranded DNA. It acts
CC progressively in a 5'- to 3'-direction, releasing 5'-
CC phosphomononucleotides.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: V01146; CAA24418.1;
CC DR PIR: A00780; NDBPT7.
CC DR PIR: S42316; S42316.
CC DR InterPro: IPR000513; Exo_N.I.
CC KM Hydrolyase; Nuclease; Exonuclease.
CC SEQUENCE 300 AA; 34502 MW; 750BE36D527E31C4 CRC64;

```

```

QY 81 VLQATILNNLEVKR-----REPETHYILSLTRDCEHFAERKEFIQFP 123
Db 60 ILEDISIKSETEKKAMAGAPIVLAFTDSVNMKKELVDPNY-----KANNR----- 104
QY 124 LSKEEVEEPPIASWVIEHEKIENFERLLRAVYAPQNIYCVHDEKSPETFEKAVKAIISCE 183
Db 105 ----AVKKRPVGY-----FE-FLDALFEREEFYCI---RBPMEGGDDVMSVIAASN 145
QY 184 PVVFIAASKLYRVYVYASWSVQADLNCMEDLLQSSVPWKYFLNTGCTDPPIKSNAMVQAL 243
Db 146 PSAFAKRAVIT-----SCDKRFTLPNCF---L 172
QY 244 KMLNGRNSMSEVPPKHETRWXYHFEVVRDTLHLTNKKDDPPYNTLM--FTGNAYIVA 301
Db 173 WCTTGNILQTE-----ESADWMLHFLQITKG--DITDGYSGIAGGDPAEDPLNPFITE 225
QY 302 SRDPYQVHLKKNPSQOLIEVVKDTYSPDEHLMATLQARMMPGSVNHPKDYDISMTSIA 361
Db 226 PRTSVLKSGKN-KGOEVIKWKVRDPEPHETLMDICKSI---GAKAGMTEEDIIKQGWMA 280
QY 362 RLVMKGHEGDDIK 375
Db 281 RLIRFNEY-NFLDK 293

RESULT 7
INRL_SHEEP STANDARD: PRT: 560 AA.
ID INRL_SHEEP
AC Q28589; Q95206;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Interferon-alpha/beta receptor alpha chain precursor (IFN-alpha-REC)
DE (interferon alpha/beta receptor-1).
GN IFNARI OR IFNAR.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Endometrium;
RX MEDLINE=97135690; PubMed=8981227;
RA Kaluz S., Fisher P.A., Kaluzova M., Sheldrick E.L., Flint A.P.F.;
RT "Structure of an ovine interferon receptor and its expression in
RL endometrium.";
RL J. Mol. Endocrinol. 17:207-215(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Endometrium;
RX MEDLINE=98006426; PubMed=9348203;
RA Han C.-S., Mathalagan N., Klemann S.W., Roberts R.M.;
RT "Molecular cloning of ovine and bovine type I interferon receptor
RT subunits from uteri, and endometrial expression of messenger
RT ribonucleic acid for ovine receptors during the estrous cycle and
RT pregnancy.";
RL Endocrinology 138:4757-4767(1997).
CC -1- FUNCTION: RECEPTOR FOR INTERFERONS ALPHA AND BETA. BINDING TO TYPE
CC I IFN TRIGGERS TYROSINE PHOSPHORYLATION OF A NUMBER OF PROTEINS
CC INCLUDING JAKS, TYK2, STAT PROTEINS AND IFN-R ALPHA-AND BETA-
CC SUBUNTS THEMSELVES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED EXCEPT
CC CONCEPTUS AT DAY 15 OF PREGNANCY.
CC -1- SIMILARITY: CONTAINS 2 FIBROBLAST TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X95939; CAA65183.1; -
 CC DR EMBL: U65978; AAB84231.1; -
 CC DR InterPro: IPR000282; CytoK_receptor_2.
 CC DR InterPro: IPR003961; FN_III.
 CC DR Pfam: PF00041; fn3; 1.
 CC KW Receptor; Transmembrane; Glycoprotein; Signal.
 CC FT SIGNAL 1 24
 CC FT CHAIN 25 560
 CC FT INTERFERON-ALPHA/BETA RECEPTOR ALPHA
 CC FT CHAIN 25 560
 CC FT EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 25 437
 CC FT DOMAIN 438 458
 CC FT DISULFID 459 560
 CC FT DISULFID 76 84
 CC FT BY SIMILARITY.
 CC FT BY SIMILARITY.
 CC FT CARBOHYD 199 220
 CC FT CARBOHYD 47 47
 CC FT CARBOHYD 55 55
 CC FT CARBOHYD 85 85
 CC FT CARBOHYD 108 108
 CC FT CARBOHYD 109 109
 CC FT CARBOHYD 112 112
 CC FT CARBOHYD 222 222
 CC FT CARBOHYD 285 285
 CC FT CARBOHYD 313 313
 CC FT CARBOHYD 359 359
 CC FT CARBOHYD 377 377
 CC FT CARBOHYD 434 434
 CC FT CARBOHYD 434 434
 CC FT CARBOHYD 352 352
 CC FT CONFLICT 522 522
 CC FT CONFLICT A -> D (IN REF. 2).
 CC FT SEQUENCE 560 AA; 63918 MW; E7198A1905D4805C CRC64;

Query Match 4.4%; Score 104.5; DB 1; Length 560;
 Best Local Similarity 22.5%; Pred. No. 1.3; Indels 65; Gaps 11;
 Matches 58; Conservative 35; Mismatches 100; Indels 65; Gaps 11;

QY 47 SOSQYCRNLY-----NFLKLPARKSINSGVTRDQDAVLQALINLEVKRR- 95
 DB 48 SSSESVRNTFSADYQILGIDNMWKLPGCHITSSKCNFSVE--LKDFEKEITELIRAE 105
 QY 96 -----PEPTDTHLSLTRDCEHFAKERTQFPLSKSEVEFPPT-----AY 135
 DB 106 EGNNTSTWTEVEFPFPLAQIOPPDVHLEAEKATILISPGTSDSITMWDLRSSFRY 165
 QY 136 SWVIEKINFERILRAVAPONICYVHDEKSE-TEKEAVKA-----IISCENVFI 188
 DB 166 SVYIMKNSSLEERTETVYPEDKIT-----KLSPEITTYCLKYKAELRLQSRVGCISPVYC 220
 QY 189 ASKLYRVVYASMSRVQADLNCMEDLLQSSVPMKYFLNTGTFPIKSNAMVQALMLNG 248
 DB 221 INTERHKVPSPEVNOIVNDQAYVLK-----W-----DPRYSTTFQAWMLR----- 263
 QY 249 RNSMESEVPKHKETRWK 266
 DB 264 --AFLKIKGKH-SNKKW 278

RESULT 8
 EXRN_BP73 STANDARD: PRT; 302 AA.
 AC P20321;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Exodeoxyribonuclease (EC 3.1.11.3) (Exonuclease).
 GN 6.
 OS Bacteriophage T3.
 OC Viruses: dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
 OC T7-like phages
 OC NCB1_Taxid=10759;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LURIA;

RX MEDLINE-90133923; PubMed-2614843;
 RA Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
 RT "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9."
 RL J. Mol. Biol. 210:687-701(1989).
 CC -1- FUNCTION: This enzyme is essential for phage DNA replication; it
 CC is believed to function in the removal of DNA-linked RNA primers.
 CC It is also necessary for host DNA degradation and phage genetic
 CC recombination.
 CC -1- CATALYTIC ACTIVITY: Degradation of double-stranded DNA. It acts
 CC progressively in a 5' to 3'-direction, releasing 5'-
 CC phosphomononucleotides.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X17255; CAA35147.1; -
 DR PIR: S07503; S07503.
 DR InterPro: IPR00513; Exo_N_I.
 KW Hydrolyase; Nuclease; Exonuclease.
 SQ SEQUENCE 302 AA; 34627 MW; D89C9CE2CC5D4773 CRC64;

Query Match 4.3%; Score 101.5; DB 1; Length 302;
 Best Local Similarity 19.2%; Pred. No. 0.99;
 Matches 61; Conservative 51; Mismatches 100; Indels 105; Gaps 15;

QY 81 VLOALINMLEVKK-----PEPTDTHLSLTRDCEHFAKERTQFP 123
 DB 60 ILEDSTKSTETRRKANVAGPIVLAFDSYNWKELVDPNT-----KANRK----- 104
 QY 124 LSKEVEFPPIAYSMVIEKINFERILRAVAPONICYVHDEKSEPTKEAVKALISCF 183
 DB 105 ----ATKKRPVGY-----FE-FLKALFEEREYCY----REPLMGDDVMGVIAIN 145
 QY 184 PNVFATSKLYRVVYASMSRVQADLNCMEDLLQSSVPMKYFLNTGTFPIKSNAMVQAL 243
 DB 146 PSAFGARKAVII-----SCQDFFTPIPCDFLWC- 174
 QY 244 KMLGNRSMSESEVPKHKETR-WKYHFEVYRDTLHTNKKDPPENLTM--FTGNAYIV 300
 DB 175 -----TTGNLTLYQTKETADWMLFQTIKG--DMTDGTSGIPGMDTAEGFLNDPFIY 224
 QY 301 ASRDFVQVHLKKNPKS--QQDLIEWKDYSPDEHLMTATLQARWMPGSGVPHRKYDISDMT 358
 DB 225 ---EPVESVLKSGKNKGQVTKWVKRAPATETLMDCINSI---GAKGMTEQELIKG 277
 QY 359 STARLYKMQGHGSDIDK 375
 DB 278 QMARILFREY-NYIDK 293

RESULT 9
 FAFY_HUMAN STANDARD: PRT; 2555 AA.
 ID FAFY_HUMAN
 AC 000507; O14601;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ubiquitin carboxyl-terminal hydrolase FAF-Y (EC 3.1.2.15)
 DE (Ubiquitin thiolesterase FAF-Y) (Ubiquitin-specific protease
 DE protease FAF-Y) (Deubiquitinating enzyme FAF-Y) (Pat facets protein
 DE related, Y-linked) (Ubiquitin-specific protease 9, Y chromosome).
 GN USPYR OR USP10 OR DEPRY.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RX NCBI Taxid=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE=97081753; PubMed=8922996;
 RA Jones M.H., Furlong R.A., Burkin H., Chalmers I.J., Brown G.M.,
 RA Khwaja O., Affara N.A.;
 RT "The Drosophila developmental gene fat facets has a human homologue
 RT in Xp11.4 which escapes X-inactivation and has related sequences on
 RT Yq11.2.";
 RL Hum. Mol. Genet. 5:1695-1701(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brown G.M., Sargent C.A., Erickson R.P., Mazayrat S.,
 RA Mitchell M.J., Jones M.H., Cooke H., Affara N.A.;
 RN Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98022381; PubMed=9381176;
 RA Lahn B.T., Page D.C.;
 RT "Functional coherence of the human Y chromosome.";
 RL Science 278:675-680(1997).
 RL [4]
 RP SEQUENCE FROM N.A., AND GENE MAPPING.
 RC TISSUE-Testis, Fetal brain, and Retina;
 RX MEDLINE=98046012; PubMed=9384609;
 RA Brown G.M., Furlong R.A., Sargent C.A., Erickson R.P., Longepied G.,
 RA Mitchell M., Jones M.H., Hargreave T.B., Cooke H.J., Affara N.A.;
 RT "Characterisation of the coding sequence and fine mapping of the human
 RT Sxb1 gene and comparative expression analysis and mapping to the
 RT Sxb1 interval of the mouse Y chromosome of the Dfry gene.";
 RL Hum. Mol. Genet. 7:97-107(1998).
 CC -1- FUNCTION: MAY FUNCTION AS AN UBIQUITIN-PROTEIN OR POLYUBIQUITIN
 CC HYDROLASE INVOLVED BOTH IN THE PROCESSING OF UBIQUITIN PRECURSORS
 CC AND OF UBIQUITINATED PROTEINS. MAY THEREFORE PLAY AN IMPORTANT
 CC ROLE REGULATORY ROLE AT THE LEVEL OF PROTEIN TURNOVER BY
 CC PREVENTING DEGRADATION OF PROTEINS THROUGH THE REMOVAL OF
 CC CONJUGATED UBIQUITIN.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN EMBRYONIC AND ADULT
 CC TISSUES.
 CC -1- DISEASE: DEFECTS IN USP9Y MAY BE THE CAUSE OF MALE INFERTILITY OF
 CC THE AZOOSPERMIA FACTOR A (AZFA) SPERMATOGENIC PHENOTYPE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
 CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y13618; CAA73940.1; -
 DR EMBL: Y13619; CAA73941.1; -
 DR EMBL: AF000986; AAC51833.1; -
 DR MEROPS: C19.028; -
 DR MIM: 400005; -
 DR InterPro: IPR001394; UCH-2.
 DR Pfam: PF00442; UCH-1; 1.
 DR Pfam: PF00443; UCH-2; 1.
 DR PROSITE: PS00972; UCH_2_1; 1.
 DR PROSITE: PS00973; UCH_2_2; 1.
 DR PROSITE: PS50235; UCH_2_3; 1.
 KW Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family;
 KW Alternative splicing.
 FT ACT_SITE 1568 1568 BY SIMILARITY.
 FT ACT_SITE 1873 1873 BY SIMILARITY.

FT ACT_SITE 1881 1881 BY SIMILARITY.
 FT VARSPIC 2071 2555 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 206 206 E -> D (IN REF. 3).
 FT CONFLICT 282 282 I -> M (IN REF. 3).
 FT CONFLICT 524 524 D -> Y (IN REF. 3).
 FT CONFLICT 542 542 S -> A (IN REF. 3).
 FT CONFLICT 666 666 L -> V (IN REF. 3).
 FT CONFLICT 883 883 C -> R (IN REF. 3).
 FT CONFLICT 907 907 F -> S (IN REF. 3).
 FT CONFLICT 946 946 N -> D (IN REF. 3).
 FT CONFLICT 1002 1002 K -> E (IN REF. 3).
 FT CONFLICT 1017 1017 K -> R (IN REF. 3).
 FT CONFLICT 1025 1027 FAN -> VAD (IN REF. 3).
 FT CONFLICT 1038 1039 KN -> RD (IN REF. 3).
 FT CONFLICT 1060 1060 T -> A (IN REF. 3).
 FT CONFLICT 1067 1067 N -> K (IN REF. 3).
 FT CONFLICT 1258 1258 F -> V (IN REF. 3).
 FT CONFLICT 1733 1733 F -> Y (IN REF. 3).
 FT CONFLICT 1953 1953 P -> L (IN REF. 3).
 FT CONFLICT 2086 2086 R -> G (IN REF. 3).
 SO SEQUENCE 2555 AA; 291115 MW; 8DEC03DF0472B995 CRC64;
 Query Match 4.3%; Score 100.5; DB 1; Length 2555;
 Best Local Similarity 19.4%; Pred. No. 19;
 Matches 102; Conservative 72; Mismatches 189; Indels 163; Gaps 24;
 QY 18 CYMLATYALKIS-----FRKCDSHLGLSRESOYCR----- 53
 DB 1297 GFALPFLDALDSKSKAMQFIIDLLHLPSTVQRLAOEOPFLMCRMGHRLPEFI 1356
 QY 54 NLVNFELPKAKRSNGSVTRGDQAVLYQTLN-----NLEV-----KKK 94
 DB 1357 TLFFTLIGSTARE-----KQYSGDYFTLRHLNLAAYGNINIPAAVLYLSEIMLKRI 1412
 QY 95 R-----EPTDTHVLSLRDCEHFAERKRFLOPLSKREVEPIYASVNIHER 142
 DB 1413 RNVKANTGETVEEPILEGH-LGVTKELLAFQTSKKYHFGCEK-----GANLIMKEL 1464
 QY 143 IENFERLRAYAPONTICVHDEKSPETFKAVYAIISCFNVTIAS---KLRYVYAS 199
 DB 1465 IDDF-----TFPASKYLYQL--RSGEL--PAEOAIPVCSSPVITINGFELVALATGC 1514
 QY 200 WSRVQADNCMDL-----LOSSVPWKYF-----LNTGDFPIKSA 237
 DB 1515 VNLKQIYDCLTEMYMGTAITTCALTEWEYLPVPGPRPPRGVGLKAGATCTYWN-- 1572
 QY 238 ENVOALKMLNG-RNSM-----ESEVPKHKETRMKRYHEVYARD 274
 DB 1573 -VIQQLYMIPIRSISILAEIGTGSDLHDMFGDEKQDSNVDPDQVDFGYPHQF---D 1628
 QY 275 TLHLTNKKKDPPEYVLMFTGNATYVASRQVHVLKPKNSQQLIEWYKDTYSPEDHMA 334
 DB 1629 KPAL-SKTEDEKREYVIGVLR-----HLOVFEGLHASQQLYYVPRGFMKQFRLMG 1677
 QY 335 TLGRKRMWPGSPVNPBKVDISMTSIRLVYKQGHGSDIK--GAPYAP---CSGIHQRA 389
 DB 1678 EPNVLRQGHDLA---EPPNSLVDSLEALKALGPAISKVLGSGFADQKICQCPHKE 1733
 QY 390 ICVYGAGDLNMMALQNHHLANKFDEKVDNMALOCLEELYRYKAIYG 435
 DB 1734 BCESEFTLLNDVIRNH-----QNLDSLEQYIKGDLLEG 1767
 RESULT 10
 TRNL_CANAL
 ID TRNL_CANAL STANDARD: PRT; 832 AA.
 AC P43075;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE tRNA ligase (EC 6.5.1.3).
 GN TRNL OR RLG1 OR LIG1.

Db 591 OFINASTNOSPDEONIVLA-TLMAMGTINIGLTMAENTPGISYROMANA----- 640
 OY 261 KETRMK-YHFEVVRDTLHETNKKDDPPYVNLMTFGNATIVASRPEVO-----HYLK 311
 Db 641 ---SQRMTDDANVRASOTILYNFOKE---OKLSYWGDC-TTSSSDGMRLSTIVRSIHADS 694
 OY 312 ND-----KSQOLIEWKDTYSPDEHLMATLQARWMPGSPVNPMPKYDISDMSIAR---- 362
 Db 695 NPHYCGKGGCTIYREFVSDLS-----AVHVAVITTNARDAHL 731
 OY 363 -LYKMGHGGDIDKGAIPAPCSGTHORALICYGAGDLNMLNHLANKFPKVDNAL 421
 Db 732 VLDGLHETDKEIEHYDTAGTYDOYFAL-----THLGFRRFAPRIIDLAD 779
 OY 422 QCL-----EYLRKAI 433
 Db 780 TKLFSTPGGEYENVAL 797

RESULT 12 MSS4_YEAST

ID MSS4_YEAST STANDARD: PRT: 779 AA.

AC P38994;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable phosphatidylinositol-4-phosphate 5-kinase MSS4 (EC 2.7.1.68)
 DE (1-phosphatidylinositol-4-phosphate kinase) (PIPSK) (PtdIns(4)P-5-
 kinase) (Dlphosphoinositide kinase).
 GN MSS4 OR YDR208W OR YD8142A.05.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.

NCBI_TaxID=4932;

SEQUENCE FROM N.A.

STRAIN=S288C;

MEDLINE=94203175; PubMed=8152413;

Yoshida S., Ohya Y., Nakano A., Anraku Y.;

"Genetic interactions among genes involved in the STT4-PKC1 pathway

Mol. Gen. Genet. 242:631-640(1994).

SEQUENCE FROM N.A.

STRAIN=S288C / AB972;

Oliver K., Shore L., Harris D., Barrell B.G., Rajandream M.A.,

Walsh S.V.;

Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.

-i- FUNCTION: CATALYSES THE PHOSPHORYLATION OF PHOSPHATIDYLINOSITOL-4-

PHOSPHATE ON THE FIFTH HYDROXYL OF THE MYO-INOSITOL RING, TO FORM

PHOSPHATIDYLINOSITOL-4,5-BIPHOSPHATE (BY SIMILARITY).

-i- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol 4-

monophosphate -> ADP + 1-phosphatidyl-1D-myo-inositol 4,5-

bisphosphate.

-i- PATHWAY: ACTS DOWNSTREAM OF STT4, BUT IN A PATHWAY THAT DOES NOT

INVOLVE PKC1.

-i- SIMILARITY: BELONGS TO THE PTINSG(4)P-5-KINASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL, D13716; BAA02869.1;

EMBL, Z68194; CAA92347.1;

PIR: S43271; S43271.

SGD: S0002616; MSS4.

InterPro: IPR002498; PIP5K.

PIfam: PF01504; PIP5K; 1.

SMART; SM00330; PIPKc; 1.

KW Transferase; Kinase.
 FT CONFLICT 610 610 F -> L (IN REF. 1).
 SQ SEQUENCE 779 AA; 89320 MW; F66B79622C9F8F6 CRC64;

Query Match 4.1%; Score 96.5; DB 1; Length 779;
 Best Local Similarity 18.3%; Pred. No. 8.2;
 Matches 76; Conservative 60; Mismatches 155; Indels 125; Gaps 16;

OY 28 KLSFLKCDSDHGLGLESRSOSQYCRN---ILNPLKLPARKSINSQVTRDQEAIVLA 84
 Db 293 KSTSLPNDNDHYVYHSGRTGNHNAANAANNNTTTTAAAGLKRSASATAEIKMKROS 352
 OY 85 ILNNLEVKRRREPFTDTHYLSLRDCEHFKARKFIOFLEKEVEFPYASGVJHEKTE 144
 Db 353 LHKREMKRRKRTF-----LVDD-----DRVLGNKVSCHVNFILAYNMVLCIRVA 399
 OY 145 -----NEERLLRAYV-----APQNYCVHDEKSPETFEKAVKAIISCF 183
 Db 400 VERCSGIMKPLPADFRFRKLAFDYHGNELTPSSQYAFKFRKYDCEVERE-LRALFGID 458
 OY 184 PNVFLASKLVRYVYASMSKVQADLNC---MEDLLQSSVPMKFFLNPCGDFPIKSAENV 240
 Db 459 PADYLVLSRSXYL-----SELNSPGKSGSEFFYSRDYKIYIKTI----- 498
 OY 241 QALKMLNGNSMSEVYPRKHKETRMKYHFEVVRDTLHETNKKDDPPYVNLMTFGNAYIV 300
 Db 499 -----HHSF-----HHLKHKHIDYNNHNRDNTLQCFYGLHRYK 535
 OY 301 ASRDFVQVHLKNPKSOQLIEWKDTYSP--DEHLMATLQARWMPGSPVNPMPKYDISMT 358
 Db 536 MPISF-QNKRIRKRIYFIV--NMNLEPPPHLDIHITYDLKGSYV--GRFYNLKERLAKDR 590
 OY 359 STARLYKMGHGGDIDKGAIPAPCSGTHORALICYAGDLNMLNHLANKFDP 414
 Db 591 STRPYMK-----DLNMLEBQKI--KFGP 612

RESULT 13 UL87_HSVSA

ID UL87_HSVSA STANDARD: PRT: 731 AA.

AC 001007;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Gene 24 protein.

GN 24.

OS Herpesvirus saimiri (strain 11).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Rhadinovirus.

NCBI_TaxID=10383;

SEQUENCE FROM N.A.

MEDLINE=92333688; PubMed=1321287;

Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,

Neman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,

Honess R.W.;

"Primary structure of the herpesvirus saimiri genome.";

J. Virol. 66:5047-5058(1992).

-i- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BCRF1,

HSV-6 US8, HVS-1 24 AND HCMV UL87.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL, X64346; CAA45647.1;

PIR: D36808; D36808.

InterPro: IPR004285; UL87.

DR Pfam; PF03043; UL87; 1.
SQ SEQUENCE 731 AA; 83572 MW; B6B35F2A8EEBF3D3 CRC64;

Query Match	4.0%;	Score 94.5;	DB 1;	Length 731;
Best Local Similarity	22.2%;	Pred. No. 11;		
Matches 81;	Conservative 43;	Mismatches 130;	Indels 111;	Gaps 21.

QY	6	RLCOLHYMALGCMYMLLATVATLKLSF--RLKCDSPHGLSRESOSQCRNTVNFLLKPLA	64
		: : : : : : : : : : : : : : : : : : : :	
Db	24	RDCNFNFQJLKFECCLAINQVYSSIDFQDLOCIMH---ESTFYTCRAVRKLLIGSNMYFF	80
QY	65	KRSINCSGV--TRGDQEAVALAILNNLEVYKKKREFFTDHYLSLTRDCEHNEKAEKRFIQP	123
		: : : : : : : : : : : : :	
Db	81	IDLTINESATGTRGMVNGPGLIINN-----TDSY- KILNICSN-----KULPIV	124
QY	124	LSKEVEPPIAVSMVIEHEKINERELLRAVYAPONITCVAHDEKSPETFEAKAITSCF	163
		: : : : : : : : : : : : : :	
Db	125	YSLFTTMDP-----HEPL-----AYRAIFYPDL-----EQTPIDYMKMFKICR--	163
QY	184	PNVFIASLAVRVVYVYASMSRVQADLNCMEDLQSSVPPKRYFLNTCGTFFPIKSNAMEQAL	243
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	164	-----YVLTSELE-----ECYEYFATVSP-----PRVNTCK-----KNYLRLVSAL	200
QY	244	KMLGNRSMSESEVPDKHKEITRMKHYHFEVVR-----DTLHLITKKK-----DPEPYNL	290
		: : : : : : : : : : : : : : : : : : : : : : : : :	
Db	201	KTLPS--TYLAIIPNPPDQLEF---EKFSLISMQEMSLNSLNTTKKIIIAVYHSHPIVI	266
QY	291	TMTFGNAV-----IVASRDVFQVHLKN-----PKSOQL-----IEWVK	323
		: : : : : : : : : : : : : : :	
Db	257	KLCSQNAEKEIKITDANFIEMQQITVNHVMPFHAETIKRDPGSRPLKVSVLLPDGVKVV-	315
QY	324	DTYSP 328	
		: : : : : : : : : : : : : :	
Db	316	-IYPP 319	

ID	THBG_BOVIN	RESULT 14
AC	09TTC36;	STANDARD; PRT; 411 AA.
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Thyroxine-binding globulin precursor (T4-binding globulin).	
GN	SERPINA7 OR TBG.	
OS	Bos taurus (Bovine).	
OC	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC	Bovidae; Bovinae; Bos.	
OX	NCBI_TaxID:9913;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Janssen O.E., Lahner H., Grasberger H., Spring S.A., Einspänner R.;	
RT	"Characterization and primary structures of bovine and porcine	
RT	thyroxine-binding globulin.";	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
CC	-1- FUNCTION: MAJOR THYROID HORMONE TRANSPORT PROTEIN IN SERUM (BY	
CC	SIMILARITY).	
CC	-1- SUBCELLULAR LOCATION: Extracellular.	
CC	-1- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER; FOUND IN PLASMA.	
CC	-1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
EMBL	AF204928; AAF15301.1; -	
OR	HSSP; P01011.2;ZACH.	

DR InterPro; IPR000295; Leuserpin2
DR InterPro; IPR000215; Serpin.

DR SMART: SM00093; SERPIN: 1.
DR PROSITE: PS00284; SERPIN: 1.
KW Serpin; Glycoprotein; Plasma; Signal.

[illegible]

RESULT	15				
RA50_PyRHO					
ID	RA50_PyRHO	STANDARD:	PRT:	879	AA.
AC	O56887;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	DNA double-strand break repair rad50 Arpase.				
GN	RAD50 OR PH0929.				
OS	Pyrococcus horikoshii.				
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.				
OX	NCBI_TaxID=539553;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=OT3;				
RX	MEDLINE=98344137; PubMed=9679194;				
RA	Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,				
RA	Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,				
RA	Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,				
RA	Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,				
RA	Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,				
RA	Masuchi Y., Shizuya H., Kikuchi H.;				
RT	*Complete sequence and gene organization of the genome of a hyper-				
RL	thermophilic archaeobacterium, Pyrococcus horikoshii OT3.*;				
CC	DNA Res. 5:55-76(1998).				
CC	-I- FUNCTION: Involved in DNA double-strand break repair (DSBR). The				
CC	rad50/mre11 complex possesses single-strand endonuclease activity				
CC	and ATP-dependent double-strand-specific exonuclease activity.				

DNA REPAIR INVOLVED IN DNA double-strand break repair (DSBR). The rad50/mei1 complex possesses single-strand endonuclease activity and ATP-dependent double-strand-specific exonuclease activity.

```

CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000004; BAA30025.1;
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR001238; RecF.
DR InterPro: IPR003405; SMC_C.
DR Pfam: PF00470; RecF; 1.
DR Pfam: PF02483; SMC_C; 1.
DR DNA_repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (By SIMILARITY).
FT DOMAIN 141 744 COILED COIL (POTENTIAL).
SQ SEQUENCE 879 AA; 103673 MW; FD4E30FD1BDCB29 CRC64;

Query Match 4.0%; Score 93.5; DB 1; Length 879;
Best Local Similarity 20.1%; Pred. No. 16; Mismatches 107; Indels 109; Gaps 17;
Matches 70; Conservative 62;

QY 19 YMLATVALKLSFRLKCDSDHGLGRESOSQYCHNLLYNFLKLPARKSINCQVT 74
DB 84 YKVLDFARNVSYLKRDLGRMRHVTETSMESVSFIDRIIPYVF-----LNATYVR 136
QY 75 RGDQEAIVLQA-----ILNNLEVKKKRREPTDTHYLSLTRDCEHFKAERKFIQRP 124
DB 137 OGQIDAILSEDETDRKIYKEILNLDLKEKA-----YDNIGKIRKIKYKSI 181
QY 125 SKEVEEPPIAASVMEHEKIENPERLLRAVVAQNTYCVHVEDEKSPETFEAVKATISCF 184
DB 182 EEKE-----KFTM-----KTENIEDLIR-----TOEKS-----FTEVLNEI----- 212
QY 185 NVFLASKLVRYVYASWSRHOADLNCMEDLQSSVPKRYFLNTCGTDFPIKSNAEWYQAL 244
DB 213 -----RMISSTWLPRLRRELBGIRKEVKT-----LEATFNSITTELKRLG 251
QY 245 MLNGR-----NSMSESEVPKPKHKEKTRWKYHFEVYRDTLHLTNKKKDDPPYNTLMTGN 296
DB 252 ELNGKRGRLERIRIOLGEGIEKKRKS--KELEEVYKELPELEKKETE--YRLIEFKD 306
QY 297 AYIVASRDFVQH--VLKNPKSQQLIEWVKDTPSPDEHLMATILQRRKRM 342
DB 307 EYLAKRNELEKRLGILSN-RLOEVKRRKIKDAES-----KVARIIRWI 346

```

Search completed: September 26, 2002, 09:51:24
Job time: 373 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 09:44:16 ; Search time 61.88 Seconds
(without alignments)
1224.495 Million cell updates/sec

Title: US-09-874-390-2
Perfect score: 2360
Sequence: 1 MYQMKRLCOLHYLMALGCMY.....NALQCLEXYLRKAYITTEL 438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2360	100.0	438	4	095395
2	1949	82.6	440	12	099CM3
3	1939	82.2	440	12	091K2
4	1728.5	73.2	398	11	09DBA3
5	1107.5	46.9	428	11	064165
6	868	36.8	453	4	09PI09
7	715	30.3	356	11	09D2A8
8	531.5	22.5	314	4	09HCY8
9	379.5	16.1	412	5	0962C3
10	379	16.1	472	5	020406
11	373	15.8	420	5	0962C5
12	371	15.7	445	5	019730
13	367	15.6	401	5	002314
14	360.5	15.3	436	5	019729
15	352	14.9	425	5	002234
16	349.5	14.8	402	5	002315

17	336.5	14.3	352	5	0962C1	0962C1 caenorhabdi
18	287.5	12.2	283	5	0962C4	0962C4 caenorhabdi
19	271	11.5	470	5	016776	016776 caenorhabdi
20	267	11.3	467	5	09U3H9	09U3H9 caenorhabdi
21	259.5	11.0	322	16	09A4H4	09A4H4 caulobacter
22	254.5	10.8	189	5	0962C6	0962C6 caenorhabdi
23	254	10.8	454	5	045866	045866 caenorhabdi
24	253.5	10.7	513	5	09TVP0	09TVP0 caenorhabdi
25	253.5	10.7	876	5	09W034	09W034 drosophila
26	248.5	10.5	513	5	09XTM4	09XTM4 caenorhabdi
27	232	9.8	489	5	021796	021796 caenorhabdi
28	219	9.3	459	5	002309	002309 caenorhabdi
29	215	9.1	864	11	09EP10	09EP10 rattus norv
30	213.5	9.0	486	5	093643	093643 caenorhabdi
31	212.5	9.0	367	5	022481	022481 caenorhabdi
32	210.5	8.9	478	5	09XX12	09XX12 caenorhabdi
33	206	8.7	865	4	09H1B5	09H1B5 homo sapien
34	205.5	8.7	753	5	002312	002312 caenorhabdi
35	205	8.7	616	11	09EP10	09EP10 mus musculu
36	199	8.4	472	5	076634	076634 caenorhabdi
37	198.5	8.4	789	11	09EP11	09EP11 mus musculu
38	197	8.3	827	4	09H1B6	09H1B6 homo sapien
39	195	8.3	821	11	09EP11	09EP11 rattus norv
40	183	7.8	424	10	09LE60	09LE60 arabidopsis
41	183	7.8	424	10	094A75	094A75 arabidopsis
42	177	7.5	406	10	09LNM5	09LNM5 arabidopsis
43	176	7.5	448	10	09Z0Z7	09Z0Z7 arabidopsis
44	166	7.0	395	10	09C9A1	09C9A1 arabidopsis
45	164.5	7.0	447	10	09LR71	09LR71 arabidopsis

ALIGNMENTS

RESULT ID	1	095395	PRELIMINARY:	PRT:	438 AA.
AC	095395:				
DT	01-MAY-1999 (TREMBLrel. 10, Created)				
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE (N-ACETYL)				
DE	TRANSFERASE 3, MUCIN TYPE).				
GN	C2/4GNT.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-BRAIN:				
RX	MEDLINE=99115671; PubMed=9915862;				
RA	Yeh J.-C., Ong E., Fukuda M.;				
RT	"Molecular cloning and expression of a novel beta-1, 6-N-				
RT	acetylglucosaminyltransferase that forms core 2, core 4, and I				
RT	branches.";				
RL	J Biol. Chem. 274:3215-3221(1999).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99143102; PubMed=9988682;				
RA	Schwentek T., Nomoto M., Levery S.B., Merx G., van Kessel A.G.,				
RA	Bennett E.P., Hollingsworth M.A., Clausen H.;				
RT	"Control of O-glycan branch formation. Molecular cloning of human CDNA				
RT	encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core				
RT	2 and core 4.";				
RL	J. Biol. Chem. 274:4504-4512(1999).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=COLON ADENOCARCINOMA;				
RA	Strausberg R.;				
DR	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF102542; AAD10824.1; -				
DR	EMBL; AF038650; AAD21525.1; -				

DR EMBL: BC017032; AAH17032.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 DR Transferrase; Glycosyltransferase.
 KW SEQUENCE 438 AA; 50863 MW; 1F0A7B451C8407.CRC64;

Query Match 100.0%; Score 2360; DB 4; Length 438;
 Best Local Similarity 100.0%; Pred. No. 3.5e-189;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVOMKRLCOLHYLMALGCMYLAATVALKLSFRLKCDSDHGLGLESRESOSQYCRNIIYNFL 60
 DB 1 MVOMKRLCOLHYLMALGCMYLAATVALKLSFRLKCDSDHGLGLESRESOSQYCRNIIYNFL 60
 OY 61 KLPARKSINCSGVTGRDQEAVALQAILNNLEVKKKREPFTDTHYLSLRDCEHFKAEKRFI 120
 DB 61 KLPARKSINCSGVTGRDQEAVALQAILNNLEVKKKREPFTDTHYLSLRDCEHFKAEKRFI 120
 OY 121 QFPLSKEEVEFPPIAYSMVHEKIEENFERLLRAYAPONITCVHDEKSPETFEKAVKAI 180
 DB 121 QFPLSKEEVEFPPIAYSMVHEKIEENFERLLRAYAPONITCVHDEKSPETFEKAVKAI 180
 OY 181 SCFENPVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKRYFLNTGCTDPPIKSNAMV 240
 DB 181 SCFENPVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKRYFLNTGCTDPPIKSNAMV 240
 OY 241 QALKMLNGRNSMESEVPKHKETRMKYHFEVYRDTHLTKKKKDPPEYNTLMTGTNAIY 300
 DB 241 QALKMLNGRNSMESEVPKHKETRMKYHFEVYRDTHLTKKKKDPPEYNTLMTGTNAIY 300
 OY 301 ASRDEVOHYLKNPKSQQLIEWVKDTPSPDEHLMATLQARMPGSPVNHKPYDISDPTS 360
 DB 301 ASRDEVOHYLKNPKSQQLIEWVKDTPSPDEHLMATLQARMPGSPVNHKPYDISDPTS 360
 OY 361 ARLVKGQHGEGDIDKGAAPAPCSGIIHORAICVYGAGDLNMLQNHLLANKFDPKVDNA 420
 DB 361 ARLVKGQHGEGDIDKGAAPAPCSGIIHORAICVYGAGDLNMLQNHLLANKFDPKVDNA 420
 OY 421 LOCLEELYLRKAIYGTGL 438
 DB 421 LOCLEELYLRKAIYGTGL 438
 RESULT 2
 OY99CW3 PRELIMINARY; PRT; 440 AA.
 AC O99CW3;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DE 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE VIRAL BETA-1,6-N-ACETYLGUCOSAMINYLTRANSFERASE.
 OS Bovine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 OX NCBI_TaxID=10385;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20583805; PubMed=1152491;
 RA Zimmerman W., Broll H., Ehlers B., Buhk H.-J., Rosenthal A.,
 RA Goltz M.;
 RT "Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
 RT Identification of an Origin of DNA Replication.";
 RL J. Virol. 75:1186-1194(2001).
 DR EMBL: AF318573; AK07999.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KW Transferrase; Glycosyltransferase.
 SEQUENCE 440 AA; 50763 MW; C88B3EAB8FDB8D9.CRC64;

Query Match 82.6%; Score 1949; DB 12; Length 440;
 Best Local Similarity 82.0%; Pred. No. 8.3e-155;

Matches 360; Conservative 38; Mismatches 39; Indels 2; Gaps 2;

OY 1 MVOM-KRLCOLHYLMALGCMYLAATVALKLSFRLKCDSDHGLGLESRESOSQYCRNIIYNFL 59
 DB 3 MACWKKKLCGHHLMALGCMYLAATVALKLSFRLKCDSDHGLGLESRESOSQYCRNIIYNFL 62
 OY 60 LKLPARKSINCSGVTGRDQEAVALQAILNNLEVKKKREPFTDTHYLSLRDCEHFKAEKRFI 119
 DB 63 LKLPARKSINCSGVTGRDQEAVALQAILNNLEVKKKREPFTDTHYLSLRDCEHFKAEKRFI 122
 OY 120 IOPPLSKEEVEFPPIAYSMVHEKIEENFERLLRAYAPONITCVHDEKSPETFEKAVKAI 179
 DB 123 IOPPLSKEEVEFPPIAYSMVHEKIEENFERLLRAYAPONITCVHDEKSPETFEKAVKAI 182
 OY 180 ISCFENPVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKRYFLNTGCTDPPIKSNAMV 239
 DB 183 ISCFENPVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKRYFLNTGCTDPPIKSNAMV 242
 OY 240 VOALKMLNGRNSMESEVPKHKETRMKYHFEVYRDTHLTKKKKDPPEYNTLMTGTNAIY 299
 DB 243 VOALKMLNGRNSMESEVPKHKETRMKYHFEVYRDTHLTKKKKDPPEYNTLMTGTNAIY 301
 OY 300 VASRDEVOHYLKNPKSQQLIEWVKDTPSPDEHLMATLQARMPGSPVNHKPYDISDPTS 359
 DB 302 VASRDEVOHYLKNPKSQQLIEWVKDTPSPDEHLMATLQARMPGSPVNHKPYDISDPTS 361
 OY 360 IARLVKGQHGEGDIDKGAAPAPCSGIIHORAICVYGAGDLNMLQNHLLANKFDPKVDNA 419
 DB 362 IARLVKGQHGEGDIDKGAAPAPCSGIIHORAICVYGAGDLNMLQNHLLANKFDPKVDNA 421
 OY 420 ALOCLEELYLRKAIYGTGL 438
 DB 422 ALOCLEELYLRKAIYGTGL 440

RESULT 3
 OY91ZK2 PRELIMINARY; PRT; 440 AA.
 AC O91ZK2;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE BETA-1,6-N-ACETYLGUCOSAMINYLTRANSFERASE.
 GN BOREF3-4.
 OS Bovine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 OX NCBI_TaxID=10385;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=V. TEST;
 RX MEDLINE=20283898; PubMed=10811884;
 RA Vanderplasschen A., Markine-Gorlaynoff N., Lomonte P., Suzuki M.,
 RA Hiraoka N., Yeh J.-C., Bureau F., Willems L., Thiry E., Fukuda M.,
 RA Pastoret P.-P.;
 RT "A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded
 RT by Bovine herpesvirus type 4.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5756-5761(2000).
 DR EMBL: AF231105; AAF72001.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KW Transferrase; Glycosyltransferase.
 SEQUENCE 440 AA; 50701 MW; 71B9C3B6BA4949BD.CRC64;

Query Match 82.2%; Score 1939; DB 12; Length 440;
 Best Local Similarity 82.0%; Pred. No. 5.7e-154;
 Matches 360; Conservative 36; Mismatches 41; Indels 2; Gaps 2;

OY 1 MVOM-KRLCOLHYLMALGCMYLAATVALKLSFRLKCDSDHGLGLESRESOSQYCRNIIYNFL 59
 DB 3 MACWKKKLCGHHLMALGCMYLAATVALKLSFRLKCDSDHGLGLESRESOSQYCRNIIYNFL 62

QY	60	IKPAKRSINSGTTRGDQAEVLAQILNNLEVKKKKRREFPDYULSTPDCHEFKERKF	119
Db	63	LKLPKRSINSGTTRGDQAEVLAQILNNLEVKKRRLPDYULSTPDCHEFKERKF	122
QY	120	IQFPLSRKEEYFPFPAVSMVHEKLENFERLRLRVAPQNTYCHVDEKSPETFEKAVKA	179
Db	123	IQFPLSRKEELDFPFAVSMVHEKLENFERLRLRVAPQNTYCHVDEKSPETFEKAVKA	182
QY	180	ISCPENWFIASKLVRVYVYASKRSVQADLNCMEDLLQSSVPMKYFLNTCGTDFPIKNAEM	239
Db	183	ISCPENWFMASKLVRVYVYASKRSVQADLNCMEDLLQSSVPMKYFLNTCGTDFPIKNAEM	242
QY	240	VOALKMLNGRSMSEVPPKHKERMKYHEHYERDTHLNNKKRDPDPVNLTFETGNAT	299
Db	243	VLAALKMLKGRSMSEVPPKHKERMKYHEVDTLTPYTSIKRDPDPVNLTFETGNAT	301
QY	300	VASBDFVOYHVLKKNRSQOLIEVWKDYTSPEHLMATLQARAWMPGSVPHNPKDKDISMS	359
Db	302	VASRAFYOHVLDNFKSQGLIVEWVWDYTSPEHLMATLQARAWMPGSVPHNPKDKDISMS	361
QY	360	IARLVKQGHGEGIDDKAPPAFPCSGTHQARICVYAGADLNMWLNQNHLLANKFDPKVDN	419
Db	362	IARLVKQGHGEGIDVSMAPAPPCSGTHQARICVYAGADLNMWLNQNHLLANKFDPKVDN	421
QY	420	ALQCELEYLRKATYGTGL 438	
Db	422	VLOCELEYLRKATYGTGL 440	
RESULT	4		
Q9DBA3		PRELIMINARY; PRT; 398 AA.	
AC	Q9DBA3;		
RC	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	2010013H22RIK PROTEIN.		
GN	2010013H22RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;		
RX	MEDLINE=21085660; Pubmed=11217851;		
RA	Kawai J., Shlnagawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,		
RA	Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kondo S., Yamanaoka I.,		
RA	Salto T., Okazaki Y., Gotojohri T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Glisic C., King B., Koehliwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schirral L.M., Seaudli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Guastinchin S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Rindwald C., Seta T., Shibata Y., Storch K.-F.,		
RA	Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,		
RA	Yasuhara-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,		
RA	Hayashizaki Y.;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
DR	EMBL: AK008234; BAB25548.1; -		
DR	MGD: MGI:1919327; 2010013H22rik.		
DR	InterPro: IPR003406; Branch.		
SO	Feature: 398 AA; 4600 MW; 275A8E93909CC4B6 CRC64;		
	PE02485; Branch: 1.		

```

Query Match          73.2%  Score 1728.5;  DB 11;  Length 398;
Best Local Similarity 77.9%;  Fred. No. 2e-136;
Matches 311;  Conservative 44;  Mismatches 43;  Indels 1;  Gaps 1;

OY 40 LGLERESOSQYCRNLTLYNFKLPPKRSINSGVTRGDOEAVLQALINNLVEYKKRREP 99
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MDLSEEFOSQYCRDLTKTKLPPKSSINSGVIRGEKAVTQALINNLTKKQOFLT 60
OY 100 DTHYLSTRDCEHFAERKEIQPPLSKSEVEERPIAYSMVIEKIEINFERLLRAYAPONI 159
   : | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 EADYLRPMRADCEHFFTKRRKFIQVPLSKKEASPRIAYSMVIEKIEINFERLLRAYTPQNV 120
OY 160 YGVHNDSEPFERKAVNALISCFENNVFASKLYVYVAWSRQADLNCMDLLQSSVP 219
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 YGVHNDQSSSEFKQAVRAYSCEFNVFASKLYVYVAWSRQADLNCMDLLQSSVP 180
OY 220 WKPYLNTGCTDEPIKSNEMVALMLNGSRNMESEVPPKHETPMKYHFEFVARDTLHLT 279
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 181 WKYLLNTGCTDEPIKSNEMVALMLNGSRNMESEVPPKHKSMTKHVEYT-DTLHMT 239
OY 280 NKKKQPPRYNLTMPTGNAYTASRDPVOHYLKNPSQOLIEVYKOTYSFDEHMLTLORA 339
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 240 SKRKPPRPNNLTMTFGNAYMASRDFIEHFVFSNKAROLIEVYKOTYSFDEHMLTLORA 299
OY 340 RMPSPVPHRPHTYDSDMTSIRLYKWKQGHEDIDKQAPARCSGIHQALICVYAGADIN 399
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 300 SWMPSPDPLRRKFDMSDMRAIRLIRKWDHEDIENGARTYSCGIHQRAVCVYSGDLH 359
OY 400 WMLQNHHLANKFDPKVDNDNALQCLEEYIRKAYITGTEL 438
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 360 WILQNHHLANKFDPKVDNDNYLCLEEYLRHKAITYGTEL 398
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
064165 PRELIMINARY: PRT; 428 AA.
AC 064165;
AC 01-NOV-1996 (TREMBLREL_01, Created)
DT 01-NOV-1996 (TREMBLREL_01, Last sequence update)
DT 01-JUN-2001 (TREMBLREL_17, Last annotation update)
DE ENZYMATIC GLYCOSYLATION-REGULATING GENE PROTEIN.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;
RA MEDLINE=96013636; Pubmed=7560067;
RA Nishio Y., Warren C.E., Buczek-Thomas J.A., Rulfs J., Koya D.,
RA Aiello L.P., Feener E.P., Miller T.B., Jr., Dennis J.W., King G.L.;
RA "Identification and characterization of a gene regulating enzymatic
RA glycosylation in rat cardiac tissue."
RA specifically in rat cardiac tissue."
RL J. Clin. Invest. 96:1759-1767(1995).
DR EMBL: S79797; AAB35697.2; -.
DR InterPro: IPR003406; Branch.
PI Pfam: PF02485; Branch; 1.
SQ SEQUENCE 428 AA; 49826 MW; AE25A18172897AA8 CRC64;

```

[illegible]

```

OY 186 VFASQSVVYVYASRVADLNCMEDLIQSSVPMKVFINTGCTDPPIKSNAMNOALKM 245
DB 176 VFASQSVVYVYASRVADLNCMEDLIQSSVPMKVFINTGCTDPPIKSNAMNOALKM 235
OY 246 LGRNSMESEVPPKHEFTKMYHFEVVRDTLHLTNKKDPPNLTMTGNATVSRDP 305
DB 236 FVGENSLFLEKMPKNEERKKRYTVDDGLNTGVKKQAPRLKTLFSSAATFVYTR 255
OY 306 VOHVLNPKSQQLIENVKDTYSPDEHLMATLQARMPGSPVHPKYYDISDMSIARLV 365
DB 296 VGVLELNKNYQKEMEWAOPTYSDEFLMATIQRIPEVPGSLPSSHKYDLSDMNAVARFV 355
OY 366 MOGHEDIDKGAAPCSGHOAICVYAGDMMMLQNHLLANKFEDPRVDNALQCLE 425
DB 356 MOFEBSDVNGAPYPCSGVHRSVCVFGVGLSMLRKHHPANKFDMVDVPAQLQCLE 415
OY 426 EYLRYKAI 433
DB 416 EHLRHKAL 423

```

```

RESULT 6
O9P109 PRELIMINARY; PRT; 453 AA.
AC 09P109:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CORE 2 BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE 3.
GN C2GNT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI.TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20219156; PubMed=10753916;
RA Schwenk F., Yen J.C., Levery S.B., Keck B., Merx G.,
RA van Kessel A.G., Fukuda M., Clausen H.;
RT "Control of O-glycan branch formation. Molecular cloning and
RT characterization of a novel thymus-associated core 2 beta1,6-N-
RT acetylglucosaminyltransferase";
RT J. Biol. Chem. 275:11106-11113(2000).
DR EMBL: AF132035; AAF63156.1; -.
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 453 AA; 53052 MW; B4379AD4427F41CA CRC64;

```

```

Query Match 36.8%; Score 868; DB 4; Length 453;
Best Local Similarity 43.0%; Pred. No. 2, 2e-64;
Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;
OY 13 LMAAGCYMLATYALKSLRICKDSDHLESGESOSQICRNILYNFLKIPARSRINCSG 72
DB 23 LMLSLSLKLLN-----RLFPQKDIYLVESLSTSPFRN--RYTHVAKVEAYEVNCSG 75
OY 73 VTRGDDEAVLQATILNMLEVKKR--EPPTDHYLSTRDCHEFAEKRFIOPLSKFEVER 131
DB 76 IV--EDEPL--ETGKSLERRRDIIDLEDDVYAMTSDCIYTLGLYAKLVSKKEKSF 131
OY 132 PLAYSMVIEKIEFRLRAVYAPONIVCVHDEKSPETFEKAVKAIISCFPNVFIASK 191
DB 132 PLAYSIVVHKDAIMVERLLHAIYNQNIYCIHYDRKAPDFKVAAMNNKLKCFENIFASK 191
OY 192 LVAVYVYASRVADLNCMEDLIQSSVPMKVFINTGCTDPPIKSNAMNOALKM LGRNS 251
DB 192 LEAVEVAHISRLQADNLCSLSSKSIOMKYVINTLGGDFPLKSNFELVSELKLGANM 251
OY 252 MESEVPPKHEFTKMYHFEVVR---DTLHL---TNKKDPPNLTMTGNATVSRDP 305
DB 252 LFTVKKPNSKLEKFTYHHELRVYEVVVKLPITINISKEAPRNIOIFVGSATFVLSQAF 311

```

```

OY 306 VOHVLNPKSQQLIENVKDTYSPDEHLMATLQARMPGSPVHPKYYDISDMSIARLV 365
DB 312 VKYIFNNSIYODFFAWMSKITYSPDEHFMATLIVPQIPGEI--SRSAQDVSDLSQKTRLV 370
OY 366 MOGHEDIDKGAAPCSGHOAICVYAGDMMMLQNHLLANKFEDPRVDNALQCLE 425
DB 371 WNYEGGF-----YPCSTGSHLRVSCITYGAELRWILKQGHFANKFDSKVPDILIKCLA 425
OY 426 EYL 428
DB 426 EKL 428

```

```

RESULT 7
O9D2A8 PRELIMINARY; PRT; 356 AA.
AC 09D2A8:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 5330430K10RIK PROTEIN.
GN 5330430K10RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI.TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PIUITARY GLAND;
RX MEDLINE=21085660; PubMed=11217851;
RX Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RX Aizawa K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RX Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RX Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RX Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RX Fieischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RX Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RX Schmitt L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RX Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RX Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RX Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RX Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RX Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RX Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RX Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RX Suzuki H., Toyono-Oka K., Wang K.H., Wetzler C., Whitaker C., Williams L.,
RX Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL: AK019924; BAB31918.1; -.
DR MGI: 1925531; 5330430K10RIK.
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch; 1.
SQ SEQUENCE 356 AA; 40446 MW; 0D67F48615656FD6 CRC64;

```

```

Query Match 30.3%; Score 715; DB 11; Length 356;
Best Local Similarity 46.9%; Pred. No. 9, 9e-52;
Matches 136; Conservative 48; Mismatches 90; Indels 16; Gaps 4;
OY 100 DTHYLSLTR--DCEHFAEKRFIOPLSKFEVERPLAYSMVIEKIEFRLRAVYAPON 158
DB 61 DTSPHPLERSCQYRIQSHYITSLSEDAAPFLAYIVIHKDPFTEFLRAIYMPON 120
OY 159 IYCVHDEKSPETFEKAVKAIISCFPNVFIASKLVYVYASRVADLNCMEDLIQSSV 218
DB 121 YVCVHADSRAITDFEKAVALQSLSCFPNAPLASKEVGVYVGGFSRLQADLNCMDLAVSKY 180
OY 219 PMKYFINTGCTDPPIKSNAMNOALKM LGRNSMESEVPPKHEFTKMYHFEVVR--- 275
DB 181 PMKYFINTGCTDPPIKSNAMNOALKM LGRNSMESEVPPKHEFTKMYHFEVVR--- 240

```


QY 276 -LHLTKKKDPPEYNLTMTGNAYVASRDEPVCHLKNPKSQOLIEWYKDTYSPDEHMA 334
 DB 241 FMHKTNLTKPPHQLIITGTATVALTRDPVNTLNDERALILEMSKDTYSPDEHFW 300
 QY 335 TLOARMPGSGVPMHPKXYSIDMTSIARLYKMOGEGDIDKAPYAPCSG 384
 DB 301 TLNRIIPGVGSGMPNPMASW-----TGNLRAYKW-----MDWEAKHGCGHG 339

RESULT 8
 Q9HCY8 PRELIMINARY; PRT; 314 AA.

AC 09HCY8;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE D1113D9.2 (NOVEL PROTEIN SIMILAR TO BETA
 1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE) (FRAGMENT).
 GN D1113D9.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL109806; CAC03741.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 FT NON_TER 1
 FT NON_TER 314 314
 SQ SEQUENCE 314 AA; 36042 MW; 8E1B3F93976B2EBE5 CRC64;

Query Match 22.5%; Score 531.5; DB 4; Length 314;
 Best Local Similarity 39.7%; Pred. No. 1.9e-36;
 Matches 100; Conservative 58; Mismatches 71; Indels 23; Gaps 6;

QY 123 PLSEEEVFPPIAYSMVTHEKIEKFERLLRAVYAPQNTYCVHDEKSETEFEAKAIISC 182
 DB 1 PLSEEDDSFLAYITTHKELAMFVOLLRAIYVPONYCHVDEKAPMKYXTAAVQTLVNC 60
 QY 183 FPNVFIASKLVRYVYASMSRYOADLNCMEDLQSSVPWKYFLNCTGDEPIKSAEMVQA 242
 DB 61 FENVFISKTEKVAAGTFLQADINCKMAYVHSEKPOMNYYINLCGODEPPIKTRREIHY 120
 QY 243 LK-MLNGRNSKSESVPRKH-KETRWKYHFEVY-RDTLHL--TNKKDDPPYNYLTMTGNA 297
 DB 121 IRSKMSKRNITPGVIOPLHISKTSQSHLEFVPGSITAPPNPFKDKPHNLITYGSA 180
 QY 298 YIVASRDFVCHLKNPKSQOLIEWYKDTYSPDEHMAITLOARMPGSGVPMHPKXYSIDM 357
 DB 181 YVYLRKFEVERILTDIAKMDLQMSKDIRSEPHQYVWTLNR--LKATPN----- 228
 QY 358 TSIALRYKMOGH 369
 DB 229 -----AGWECN 234

RESULT 9
 Q962C3 PRELIMINARY; PRT; 412 AA.
 AC 0962C3;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE GLY-18 (FRAGMENT).
 GN GLY-18.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-N2;
 RA Warren C.E., Krizus A., Dennis J.W.;
 RT "Complementary Expression Patterns of Six Non-Essential Caenorhabditis
 elegans Core 2/1 N-Acetylglucosaminyltransferase Homologs."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY037797; AAK94762.1; -
 FT NON_TER 1
 SQ SEQUENCE 412 AA; 47583 MW; 5F9325D7320D4AAE CRC64;

Query Match 16.1%; Score 379.5; DB 5; Length 412;
 Best Local Similarity 25.5%; Pred. No. 1.4e-23;
 Matches 97; Conservative 76; Mismatches 150; Indels 57; Gaps 8;

QY 68 INCSGVTRGDQEAVALQALNNLEVKKKRPEPTDTHVLSLTRDCEHFAERKFIQFPLSKE 127
 DB 34 INCLINIFGSKNRNQLRIVNSRSTEDK-----LTYSTRCQTLKSLPRFKVPLSPE 85
 QY 128 EVERPIAYSMVTHEKIEKFERLLRAVYAPQNTYCVHDEKSETEFEAKAIISCPRVVF 187
 DB 86 EESFPLSYGLVYKELSQVLEMLSSIVHPQNECYIAGVENSAPIFQMLKELSCFSNTH 145
 QY 188 IASKLVRYVYASMSRYOADLNCMEDLQSSVPWKYFLNCTGDEPIKSAEMVQALKMLN 247
 DB 146 F-MKRPFLDWSHETINSAYCIEFLSHKSDMKRYFQYLSGVDIPLKTNLEWQILKLN 204
 QY 248 GRNMESEVPRPKHETRWKYHFEVYRDTLHLTNKKDDPPYNYLTMTGNAYVASRDFVQ 307
 DB 205 GTANVEIK-PYQYORLNRK-----NETQSPPL---LFKSSLSLIPREAN 246
 QY 308 HVLNPKSQOLIEWYKDTYSPDEHMAITLOARMPGSGVPMHPKXYSIDMTSIARLYKWO 367
 DB 247 HLSSSIDPQQLLEFLRLTGNADIEGFWGL-----FGKNKLFEDIGSLNFKEMI 294
 QY 368 GHEGIDDKGAPYAP-----CSGIDHQAICVYAGDLMMLQNHLLAN 410
 DB 295 SYKNNVETNLTPPDGMRYYISRDQIMSKRNCHMYMKAQSCVFGIDGPRRLKSKALVAH 354
 QY 411 KEDPKVDNALQC-LEEYLR 429
 DB 355 KEYLKSEPEAYFCLKEHRR 374

RESULT 10
 Q20406 PRELIMINARY; PRT; 472 AA.
 AC Q20406;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE F44F4.6 PROTEIN (GLY-1).
 GN F44F4.6 OR GLY-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA COLES L.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 investigating biology".
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-N2;

RA Warren C.E., Krizus A., Dennis J.W.;
RT "Complementary Expression Patterns of Six Non-Essential *Caenorhabditis*
RT *elegans* Core 2/I N-Acetylglucosaminyltransferase Homologs.";
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: 237092; CA85457.1; -
DR EMBL: AY037793; AAK94758.1; -
DR EMBL: AY037793; AAK94757.1; -
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch: 1.
SQ SEQUENCE 472 AA; 55412 MW; 36A1BD96D9C6245A CRC64;

Query Match 16.1%; Score 379; DB 5; Length 472;
Best Local Similarity 25.3%; Pred. No. 1.8e-23;

Matches 114; Conservative 83; Mismatches 162; Indels 92; Gaps 17;

QY 31 FRKCDSDHGLGSERE-----SOSQYCRNLINFLKL-----PAKRS-----INCSGV 73
DB 13 FRKFECCLOLGAMVRRFKITISQFLYSFTYVYLRIGFVDYSLVPRNKLINIPLECNV 72
74 TRGDQRAVLQALINLNEVKKRREPDTHTY-----LSLRDCEHFKAERKFIQPLSK 126
DB 73 INGT-----NRKKISRARQ--WDWHEFDWEHEIFNSRNVCSITDKYFNTRIPSSR 122
QY 127 EEEVFPFIASWVIHEKIEFPERILRAVYAPQNTYCYHVEKSEPTKEAVKAIISCFPN- 185
DB 123 EEAEPYPLAGLVYKTYIVLTQMSLEFQPHFCITVDOSFNEKSYQIALPSCFPNM 182
QY 186 -FEISKVLRYVYASWSRQADLNCMEDLQSSVPRKYFLNCGTGFPIKSNAEWQALK 244
DB 183 HVFIFGP-----SOWGSGIILKNYITTCFNNILSKSKQKKYQYISGTDLPRTINEMVRIR 239
QY 245 MUNGNSMSEVPPKAKETRMKYHEFVVRDTLTLTKKKKDPPEYNLTMTGNAYIVASRD 304
DB 240 ALNG--SMNTDYST-----FEVDR-----YKMEGVLP--MPYVKSMSVVPRE 281
QY 305 FPOHVLKNKRSQOLIMVADTISPDHLWATLO-----RAR--WMPGSVPHN 349
DB 282 GADYLLSSPRVQKILKLYSTKPTIDESFWSYSPALLPVGPSIVRDLIMRKNEFKLR 341
QY 350 PKYDISDMTSA-----RLVWQGHGEGIDKGAAPYAPCSGTHORATCYGAGDMMML 402
DB 342 PPEE-RTVNSIGTSYIGRQVGMQ-----KECFGVKDPSCVGEVDIEETL 388
QY 403 QNHHLANKFDPKVDNALQCLEEYLRYKAI 433
DB 389 TRPELVAKLYLEFQPAFAFCMLKEVRRSL 419
RESULT 11
Q962C5 PRELIMINARY; PRT; 420 AA.
AC Q962C5;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE GLY-16.
GN GLY-16.
OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;
OC Rhabdilitidae; Peloderinae; *Caenorhabditis*.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RA Warren C.E., Krizus A., Dennis J.W.;
RT "Complementary Expression Patterns of Six Non-Essential *Caenorhabditis*
RT *elegans* Core 2/I N-Acetylglucosaminyltransferase Homologs.";
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY037793; AAK94760.1; -
SQ SEQUENCE 420 AA; 48325 MW; 6BDC5BE558711614 CRC64;

Query Match 15.8%; Score 373; DB 5; Length 420;
Best Local Similarity 26.6%; Pred. No. 4.9e-23;
Matches 105; Conservative 77; Mismatches 158; Indels 54; Gaps 11;

QY 57 YNPLKLPARK-----SINCSGVRDQEAVALIINLNEVKKRREPDTHTYLSLT 107
DB 22 YLFVAKPSRPPDLDDLEINCTNLIQGLKNEKQLELINTKMTIEKLMNSIDR----- 75
QY 108 RCEHFKAERKFIQPLSKSEEEVFPFIASWVIHEKIEFPERILRAVYAPQNTYCYHVEK 167
DB 76 -CHTVKSMFRFNEVPVLSSEEAFAFLSYGLVYKELSYQLFMLSIIYQONEXCIAVGN 133
QY 168 SEPTKEAVKAIISCFPNVFIASKLYRVYVYASWSRQADLNCMEDLQSSVPRKYFLNLC 227
DB 134 SASTFILLLEESDCEPNVHF--MKRPITWGSYEIINSYVDDICKPLSHKSNMKYQYLS 192
QY 228 GNDPEPIKSNAEWQALKMLNGRSMSEVPPKAKETRMKYHEFVVRDTLTLTKKKDPP 287
DB 193 GVDIPKLTNLEMYRLKLSGTANVAETIKYENR-----LLGONETESPLP 238
QY 288 YNLTMTGNAYIVASRDFQVHLKNPKSQOLIMVADTISPDHLWATLQARMM--PG 344
DB 239 ----LFKSSLSLIPKKAANYLASSIPQQLLEFLNLTWVADGEFGTLEGNKGLFDVPG 294
QY 345 SVPNPKYDISDMTSTARIK-----W-----QGHGIDKGAAPYAPCSGTHORATCYGAG 396
DB 295 SL-NPEEWYTFKEMARANLTKPTDGFYIYSRQIWFESG-----CHNMKGSGSCVFGIG 348
QY 397 DLNWMLONHHLANKFDPKVDNALQCL-LEEYLR 429
DB 349 DVSNLQAKALVANKLYLTSEPAVFLCKLEHR 382

RESULT 12
Q19730 PRELIMINARY; PRT; 445 AA.
AC Q19730;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE F22D6.11 PROTEIN.
GN F22D6.11.
OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;
OC Rhabdilitidae; Peloderinae; *Caenorhabditis*.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkenson U.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [12]
RX MEDLINE=99069613; PubMed=9851916;
RP SEQUENCE FROM N.A.
RT none;
RT "Genome sequence of the nematode *C. elegans*: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z71262; CA95817.1; -
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch: 1.
SQ SEQUENCE 445 AA; 51330 MW; 4380A86BAF05DC6D CRC64;

Query Match 15.7%; Score 371; DB 5; Length 445;
Best Local Similarity 25.5%; Pred. No. 7.7e-23;
Matches 98; Conservative 76; Mismatches 149; Indels 62; Gaps 9;

QY 68 INCSGVRDQEAVALIINLNEVKKRREPDTHTYLSLRDCEHFKAERKFIQPLSK 127
DB 62 INCLINFGSKSNRQRIIVNSRSIEDK-----LLYSTDRQCLSLFRFNKVPVLSPE 113
QY 128 EEEVFPFIASWVIHEK-----IEFPERILRAVYAPQNTYCYHVEKSEPTKEAVKAIISG 182

Db 114 EESPPLSYGLLYKELSOVTIQRVLFMLSSITYPHONEYCIAVGENSAPITONLKEISNC 173
 QY 183 FPNVFIASKIVRVVYASMSRVOADLNCMEDLQSSVPKRYFLNCTGDFPIKSAEMVOA 242
 Db 174 FSNHMF-MKRPIDMSGHEIINSAYDCLFELSHLKSMDRRFQYLSGVDIPLKTLEWVOI 232
 QY 243 LKMINGNSMSEVPPKHKETRMKYHFEVVRDILHLTNKKKDPPIYMLTFTGNAVYAS 302
 Db 233 LKHLNGTANVEIK-PYOYORLGRK-----NETQSPLP-----LEKSSLSLIP 274
 QY 303 RDPVQHLKLNKPSQOLLEWVKDITYSPDEHLMATIQARARMGSPVNPBKIDISMTSIAR 362
 Db 275 REANHLSSSSIPQOLLEFLNLTGIDEGFWGTL-----FGNNKLPDIPGSLN 322
 QY 363 LVKMGHEGDDKCAPYAP-----CSGIRARICVYGAGDLMMLONH 405
 Db 323 FKEMISKNNVETMLTPTDGMRYTISRDOIMSKPNCHNWKAKSCVFGIGDVRLKSK 382
 QY 406 HLLANKFDPKVDNALOC-LEEYLR 429
 Db 383 ALVAHKRYLKSEPEAYFCLLEHRR 407

RESULT 13
 ID 002314 PRELIMINARY; PRT; 401 AA.
 AC 002314;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE F15D6.2. PROTEIN.
 GN F15D6.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA Dobson R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RA SEQUENCE FROM N.A.
 RX MEDLINE=99069613; Pubmed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; Z83125; CAB05620.1; -.
 DR InterPro; IPR003406; Branch.
 DR Pfam; PF02485; Branch; 1.
 SQ SEQUENCE 401 AA; 45608 MW; E1D3F8AD84C0873D CRC64;

Query Match 15.6%; Score 367; DB 5; Length 401;
 Best Local Similarity 25.5%; Pred. No. 1.5e-22;
 Matches 97; Conservative 74; Mismatches 143; Indels 66; Gaps 8;

QY 57 YNFKLAKR-----SINGSGVTRGDOEAVLQAILNLEVKKKRPTDTHYLSLT 107
 Db 22 YLFAKSSRRPTDLDDLELNTINILQGLKNEKOLELINTKMEIDKLWNSTDR----- 75
 QY 108 RDCHEFAERKFIQPLSKKEFPIIAYSMYIHEKIEFRLRAVYAPQIYCVHDEK 167
 Db 76 --CHTVSMFNFENEYPLSEEARFPLSYGLLYKELSOVLEMLSSITYPONEYCIAVGEN 133
 QY 168 SPETFEAKVKAIIISCFPNVFIASKIVRVVYASMSRVOADLNCMEDLQSSVPKRYFLNCTC 227
 Db 134 SASFTLLLELSDCFN-----KRPIWGSYEIINSYVDCLFSLHLSKMMKRYFQYLS 188
 QY 228 GTDPIPSNAEMVOALMLKNGNSMSEVPPKHKETRMKYHFEVVRDILHLTNKKKDPPI 287
 Db 189 GVDIPLKTLEWVOILKSLNGTANVEIKYENR-----LLGQNETESPLP 234

QY 288 YNLTFTGNAVYASRDPVOHLKLNKPSQOLLEWVKDITYSPDEHLMATIQARARM-----PG 344
 Db 235 -----LFKSSLSLIPRKAANYASSISIPQOLLEFLRMTWVADEGFWGTLFGNNKLPDPVG 290
 QY 345 SVRHPKRYDISDMTISARLVYMGQHEGDDIXGARYAPCSGIRARICVYGAGDLMMLONH 404
 Db 291 S-----LNFEHQIWFESG-----CHNHKQSCVFGIGDVSNLLQA 327
 QY 405 HLLANKFDPKVDNALOCL 424
 Db 328 KALVAHKLTYLSEPEAYFCL 347

RESULT 14
 ID 019729 PRELIMINARY; PRT; 436 AA.
 AC 019729;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE F22D6.12 OR GLY-19.
 GN F22D6.12 OR GLY-19.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RA SEQUENCE FROM N.A.
 RX MEDLINE=99069613; Pubmed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 RN [3]
 RA SEQUENCE FROM N.A.
 RC SPRATIN-2;
 RA Warren C.E., Krizus A., Dennis J.W.;
 RT "Complementary Expression Patterns of Six Non-Essential Caenorhabditis
 RT elegans Core 2/1 N-Acetylglucosaminyltransferase Homologs."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z71262; CAA95816.2; -.
 DR EMBL; AY037801; AAK94766.1; -.
 SQ SEQUENCE 436 AA; 50669 MW; 073ECE382E8B36F CRC64;

Query Match 15.3%; Score 360.5; DB 5; Length 436;
 Best Local Similarity 24.9%; Pred. No. 5.7e-22;
 Matches 102; Conservative 78; Mismatches 170; Indels 59; Gaps 9;

QY 39 HGLRESRSOQYCRNLIYNFLK-----LPAKRSINGSGVTRGDOEAVLQAILNN 88
 Db 20 YFVESLFPKROEDKNVSKOFLKSICTTASDYSLLDNMEINCSNMLKGYK-----TN 71
 QY 89 LEVAKKEPFDTHYLSITRCEHFKAERKTIQPLSKKEFPIIAYSMYIHEKIEFRLRAVYAPQIYCVHDEK 148
 Db 72 EKLDIMHLDIEQLFSCNTNCOFTLKLFRFNTPMSEAEEHFLSYGMVLYKLPOLVF 131
 QY 149 LLRAVYAPQIYCVHDEKSPETFEAKVKAIIISCFPNVFIASKIVRVVYASMSRVOADLN 208
 Db 132 LLSSTIHPONEYCIAVGENSAPITONLREYVSTCFSNVH-MKRPIDMSGHEIIDSYYD 190
 QY 209 CMEDLQSSVPKRYFLNCTGDFPIKSAEMVOALMLKNGNSMSEVPPKHKETRMKYH 268
 Db 191 CLFPLSHLETDRWFQYLSGVDIPLKTLEWVOILKHLNGTSNVEI---TNYQOAR----- 243
 QY 269 FEVVRDILHLTNKKKDPPIYMLTFTGNAVYASRDPVOHLKLNKPSQOLLEWVKDITYSP 328
 Db 244 -----LTGKNENESP--LPLFKSSLSAIIIPRKAANOIASSTARKLLEFLMTEIA 292

OY 329 DEHLMATLQRAHMGSPVNHPPKIDISMTSIALRVKMGHEGDI---DKGAPYAP--- 381
DB 293 DEGFNGTL-----FGNKDQFNISGSKDMWEYRDNONNINFPDGMSTYISRDQ 343
OY 382 -----CSGIHORAICVAGADLNMLNHLANKFDPKVDNALQCL 424
DB 344 IWDPELCKMYMKDSCVFGIDVPRLRTSKALVAHKFYLKSPFAEYFCL 392

RESULT 15

ID 002234 PRELIMINARY: PRT: 425 AA.
AC 002234,
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE HYPOTHETICAL PROTEIN C54C8.11.
CN C54C8.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; 283102; CAB05469.1; -
DR EMBL; AL032664; CAB05469.1; JOINED.
DR EMBL; 283102; CA21769.1; JOINED.
DR InterPro; IPR003406; Branch.
DR Pfam; PF02485; Branch; 1.
KW Hypothetical protein.
SQ SEQUENCE 425 AA; 49137 MW; 71A78A4D94AA079 CRC64;

Query Match

14.9%; Score 352; DB 5; Length 425;
Best Local Similarity 28.2%; Pred. No. 2.8e-21;

Matches 111; Conservative 58; Mismatches 167; Indels 58; Gaps 14;

OY 68 INCSGVTRGDQAVLQALINLNLEVKKKRPPFTDTHLS--LTRDCEHFKAKRRTIOPLS 125
DB 41 LECQKMFDDDPVSLAGALFKFD---DREITLQIILKLSQENNAQCAEFQKITGFFOEPS 97
OY 126 KEEVEFPPLAYSNVIEHEKLENFERLLRAVYAPQNIYCVHDEKSPETFEAVYKAIISCPEP 185
DB 98 OELEPPLAYGMLVHGDFFVQLSLLSAIQPONOFCLAVDGNSSVEFGLVRLMSRCYGN 157
OY 186 V--FIASKLVRYVYASMSKVQADLNCMEDLQSSVPWKYFLNTCGTFPIKSNAEVQAL 243
DB 158 IOYFTTDE--LRMGYEILTSVFOCVDTLAKLPSPDKYFOYLSGVDAPLKSNLEMIRIL 214
OY 244 KMLNGRNSMSESVPPKHKFTRWKHFEEVYRDTLHLNKKKDPPEVNLTMETGNAYIVASR 303
DB 215 KALNG--SFNAELP-----FEYR-----LNKKR-PWSSPLPIYKTSLSATFSR 256
OY 304 DFVQHLKNKPKSQQLIEWVKDTYSPDEHLMATLQ-----RARMPGSPVN---HPK 351
DB 257 KSNFMVNSEKYLEQIDFLRGTTCADESLMATIAGNPVKYKFSSELPMPGFPDAKAWIHKN 316
OY 352 YDISDMTSLARLVKMGHGDIDKG-----APYAPCSGIHORAICVYAGADLN 399
DB 317 Y---RRTRGKLGKY-GENOKIDNGYVYSRYOQYVNRAP-VKCKGYVYRLSCVFGYDLP 370
OY 400 WMLONHHLANKFDPKVDNALQCLEEYLYRKAI 433
DB 371 NLINRHELVAAHKLFTSYOPAPAFPCLVENSROKSM 404

Search completed: September 26, 2002, 09:50:50
Job time: 394 sec

Db 361 ARLVKNQHGHDIDKGAFCGSIHQRAICVYAGADLNMMLNHHLLANKPFCVDDNA 420
QY 421 LOCLEETLRKAIYGTCL 438
Db 421 LOCLEETLRKAIYGTCL 438

RESULT 2
US-07-955-041-4
Sequence 4, Application US/07955041
Patent No. 5360733

GENERAL INFORMATION:
APPLICANT: FUKUDA, MINORU
TITLE OF INVENTION: A NOVEL BETAL-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUCOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
TITLE OF INVENTION: ENZYMAIC ACTIVITY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/955,041
FILING DATE: 19921001

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATRYN

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 9294

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-535-9001

TELEFAX: 619-535-8949

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 428 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-955-041-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

QY 68 INCSGTVRGQDAVLAQLLNLEVK-KRREPFDTHLSLTRDCEHFKARKFIQFPLSK 126
Db 57 INCTKVLQGVNLEOKYKLEILTVYFKRRPWPDDYINNTSDCSSFIKRRKIYVEPLSK 116

QY 127 EEVEFPFAYSMVHEKTEJENRELLRAYAPONTYCVAVDEKSPETFEAVKAIISCPNV 186
Db 117 EEVEFPFAYSMVHEKTEJENRELLRAYAPONTYCVAVDEKSPETFEAVKAIISCPNV 176

QY 187 FIASKLRVYVYASWSRVQADLNCMEDLQSSVPMKYFLNTCGTFPIKSNAEVQALKML 246
Db 177 FVASRLESVYVYASWSRVQADLNCMEDLQSSVPMKYFLNTCGTFPIKSNAEVQALKML 236

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

Db 297 GYVLQNKIKQKLEMMADTSPDEILMATORLEPVPGSIPASHKYDLSMQAVAFVFW 356
QY 367 QGHGHDIDKGAFCGSIHQRAICVYAGADLNMMLNHHLLANKPFCVDDNA 426
Db 357 QFEGDYSKAGAPYPCGCVHVSVCIGAGDLMMLKHHLPFNKFPVDVDFALQICLDE 416
QY 427 YLRKAI 433
Db 417 HLRKAL 423

RESULT 3
US-08-227-455-4
Sequence 4, Application US/08227455
Patent No. 5624832

GENERAL INFORMATION:
APPLICANT: FUKUDA, MINORU
TITLE OF INVENTION: A NOVEL BETAL-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUCOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
TITLE OF INVENTION: ENZYMAIC ACTIVITY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,455
FILING DATE: 14-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATRYN

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 9957

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-535-9001

TELEFAX: 619-535-8949

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 428 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-227-455-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

QY 68 INCSGTVRGQDAVLAQLLNLEVK-KRREPFDTHLSLTRDCEHFKARKFIQFPLSK 126
Db 57 INCTKVLQGVNLEOKYKLEILTVYFKRRPWPDDYINNTSDCSSFIKRRKIYVEPLSK 116

QY 127 EEVEFPFAYSMVHEKTEJENRELLRAYAPONTYCVAVDEKSPETFEAVKAIISCPNV 186
Db 117 EEVEFPFAYSMVHEKTEJENRELLRAYAPONTYCVAVDEKSPETFEAVKAIISCPNV 176

QY 187 FIASKLRVYVYASWSRVQADLNCMEDLQSSVPMKYFLNTCGTFPIKSNAEVQALKML 246
Db 177 FVASRLESVYVYASWSRVQADLNCMEDLQSSVPMKYFLNTCGTFPIKSNAEVQALKML 236

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

QY 247 NGRNSMESEVPKPKHETRMKYHEEVVADTLHLNKKKDPPTNYLMTGNAYIVASDFV 306
Db 237 MGNNEIETEMPSHKEBRMKRYEVVNGKLTNGVTMLPLLEPLSGSAYEVVSSEV 296

BEST AVAILABLE COPY

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 08:15:46 : Search time 19.87 Seconds
(without alignments)
538.420 Million cell updates/sec

Title: US-09-874-390-2
Perfect score: 2360
Sequence: 1 MVQWKRQLCOLHYLWALGCMY.....NALQCLBEYLRYKAIYGTGL 438

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

cal number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2.6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2.6/prodata/2/1aa/6A_COMB.pep.*
4: /cgn2.6/prodata/2/1aa/6B_COMB.pep.*
5: /cgn2.6/prodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2.6/prodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2360	100.0	438	4	US-09-233-506-2
2	1142.5	48.4	428	1	US-07-955-041-4
3	1142.5	48.4	428	1	US-08-227-453-4
4	1142.5	48.4	428	1	US-08-472-482-4
5	1142.5	48.4	428	1	US-08-487-069-4
6	1142.5	48.4	428	1	US-09-233-506-3
7	784	33.2	400	1	US-08-118-906-14
8	784	33.2	400	1	US-08-486-196-14
9	784	33.2	400	1	US-08-488-135-14
10	784	33.2	400	2	US-08-474-065-14
11	784	33.2	400	4	US-09-233-506-4
12	447	18.9	126	1	US-08-118-906-4
13	447	18.9	126	1	US-08-486-196-4
14	447	18.9	126	1	US-08-488-135-4
15	447	18.9	126	2	US-08-474-065-4
16	426	18.1	126	1	US-08-118-906-2
17	426	18.1	126	1	US-08-486-196-2
18	426	18.1	126	1	US-08-488-135-2
19	426	18.1	126	2	US-08-474-065-2
20	278	11.8	64	4	US-09-233-506-10
21	203.5	8.6	49	4	US-09-233-506-14
22	126	5.3	33	1	US-08-118-906-8
23	126	5.3	33	1	US-08-486-196-8
24	126	5.3	33	1	US-08-488-135-8
25	126	5.3	33	2	US-08-474-065-8
26	118	5.0	33	1	US-08-118-906-6
27	118	5.0	33	1	US-08-486-196-6

28	118	5.0	33	1	US-08-488-135-6	Sequence 6, Appl1
29	118	5.0	33	2	US-08-474-065-6	Sequence 6, Appl1
30	109	4.6	22	1	US-08-118-906-12	Sequence 12, Appl
31	109	4.6	22	1	US-08-486-196-12	Sequence 12, Appl
32	109	4.6	22	1	US-08-488-135-12	Sequence 12, Appl
33	109	4.6	22	2	US-08-474-065-12	Sequence 12, Appl
34	103	4.4	316	1	US-08-597-236-12	Sequence 12, Appl
35	103	4.4	316	1	US-08-746-682A-12	Sequence 12, Appl
36	101	4.3	788	2	US-08-907-168-6	Sequence 6, Appl1
37	100.5	4.3	2555	3	US-09-058-489-36	Sequence 36, Appl1
38	94.5	4.0	253	3	US-08-929-329-9	Sequence 9, Appl1
39	94.5	4.0	619	4	US-09-153-310-42	Sequence 42, Appl
40	92.5	3.9	2289	3	US-09-051-019-2	Sequence 2, Appl1
41	91	3.9	3135	1	US-08-323-170B-2	Sequence 2, Appl1
42	91	3.9	3135	4	US-08-954-441-2	Sequence 2, Appl1
43	90.5	3.8	616	4	US-08-965-902A-2	Sequence 2, Appl1
44	90	3.8	1128	1	US-08-111-939-2	Sequence 2, Appl1
45	90	3.8	1194	4	US-09-092-508-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-233-506-2
; Sequence 2, Application US/09233506
; Patent No. 6136580
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jinn-Chern
; TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
; FILE OF INVENTION: Core 2, Core 4 and I Branches
; CURRENT APPLICATION NUMBER: P-LJ 3415
; CURRENT FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-233-506-2

Query Match	100.0%	Score 2360;	DB 4;	Length 438;
Best Local Similarity	100.0%;	Pred. No. 6.9e-230;		
Matches 438;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MVQWKRQLCOLHYLWALGCMYLLATVAKLSFRLKCDSDHLGLSRESQSQCRLNLYNFL	60	
DB	1	MVQWKRQLCOLHYLWALGCMYLLATVAKLSFRLKCDSDHLGLSRESQSQCRLNLYNFL	60	
QY	61	KLPKRSINCSGVTGDCQAVLQAILNNLEVKKKRPFTDTHYLSLTRDCEHFAERKFI	120	
DB	61	KLPKRSINCSGVTGDCQAVLQAILNNLEVKKKRPFTDTHYLSLTRDCEHFAERKFI	120	
QY	121	QPLSKEEVEFFIAYSMVTHEKIENFERLLRAYAPQNIYCVHVEKSPETFEAKVAKII	180	
DB	121	QPLSKEEVEFFIAYSMVTHEKIENFERLLRAYAPQNIYCVHVEKSPETFEAKVAKII	180	
QY	181	SCFPNVFIASKLRVYVYASWSRVQADLNCEMEDLLQSSVPWKYFLNCTGCTFPKSNAMV	240	
DB	181	SCFPNVFIASKLRVYVYASWSRVQADLNCEMEDLLQSSVPWKYFLNCTGCTFPKSNAMV	240	
QY	241	QALKMLNGRNSMESEYPPKHETRWYHFEVVRDTHLTNNKKDPPPYNTMTFTGNAYIV	300	
DB	241	QALKMLNGRNSMESEYPPKHETRWYHFEVVRDTHLTNNKKDPPPYNTMTFTGNAYIV	300	
QY	301	ASRDFVQHVHLKPNKSOQLIEWVADTYSPEHLWATLQARWMPGSPVNPHPKYDISDMSI	360	
DB	301	ASRDFVQHVHLKPNKSOQLIEWVADTYSPEHLWATLQARWMPGSPVNPHPKYDISDMSI	360	
QY	361	ARLVKQGHGEGIDKGAPYAPCSGIHQRAICVYVGADLNMLQNHLLANKFPKVDNNA	420	